Kooperationsanreize für autonome Einheiten in selbst-organisierenden Informationssystemen

Dissertation

zur Erlangung des akademischen Grades
Doktor-Ingenieur (Dr.-Ing.)

vorgelegt dem Rat der Fakultät für Mathematik und Informatik
der Friedrich-Schiller-Universität Jena

von Dipl.-Inform. Philipp Obreiter
geboren am 11.10.1976 in Karlsruhe-Durlach

Gutachter

1. Prof. Dr. Birgitta König-Ries

2. Prof. Dr. Christof Weinhardt

Tag der letzten Prüfung des Rigorosums: 15.2.2006

Tag der öffentlichen Verteidigung: 16.2.2006

Erklärung

Ich versichere hiermit wahrheitsgemäß, die Arbeit selbständig angefertigt, alle benutzten Hilfsmittel vollständig und genau angegeben und alles kenntlich gemacht zu haben, was aus Arbeiten Anderer unverändert oder mit Abänderung entnommen wurde.

Jena, 21. Februar 2006

für Sok Shee, meine liebe Frau

Kurzfassung

Für Informationssysteme haben die *Selbstorganisation* und *Autonomie* der Teilnehmer große Bedeutung gewonnen, da hier die Inbetriebnahme mit sehr geringen Kosten verbunden ist und die Betriebskosten durch Verzicht auf Infrastruktur entfallen. Ein solches System ist beispielsweise ein *Ad-hoc Netz* an einer Universität, in dem die Geräte von Studenten sich automatisiert untereinander Dienste erbringen. Aus der Sicht menschlicher Benutzer, die durch ihre Geräte am System teilnehmen, um die Dienste anderer nutzen zu können, ergibt sich aber ein großes Problem: Die Geräte (die *Einheiten* des Informationssystems) können sich untereinander betrügen, indem sie sich versprochene Dienstleistungen nicht erbringen.

Bisherige Lösungsansätze zur Vermeidung oder Bestrafung von Betrugsverhalten sind jedoch nicht einsetzbar, da sie etwa durch die Verwendung vertrauenswürdiger Dritter die Selbstorganisation oder die Autonomie der Teilnehmer verletzen. Andere Ansätze im Bereich der *verteilten Vertrauensbildung* kommen zwar mit diesen Rahmenbedingungen zurecht, versäumen es aber durch Mängel in der Bewertung von eigenen Erfahrungen und Empfehlungen anderer, Betrug effektiv zu unterbinden. Die Arbeit leistet in diesem Bereich Abhilfe, indem sie durch *typbasierte Glaubensbildung* verfügbare Informationen probabilistisch richtig berücksichtigt und *nicht-abstreitbare Beweismittel* als Grundlage von Empfehlungen einsetzt.

Der Systementwurf ist nur dann erfolgreich, wenn die Geräte von potentiell interessierten menschlichen Benutzern am System mit der Software des Systementwerfers teilnehmen. Hieraus ergeben sich zwei Maximen: Einerseits muss der Entwurf hinreichend gutartiges Verhalten in *Normen* vorschreiben. Andererseits dürfen Einheiten keinen Anreiz haben, die Software zu *manipulieren*, um sich opportunistisch zu verhalten. Dies führt zur Typisierung der Einheiten in *normative* und *strategische* Einheiten, die die originale beziehungsweise manipulierte Software verwenden. Einige *technische* und *rechtliche Hindernisse* sorgen dafür, dass die Verwendung manipulierter Versionen der Software für einen Benutzer nicht zwingend von Vorteil ist und damit der Entwurf des Systems überhaupt sinnvoll ist. Das Modell zur Glaubensbildung baut auf der Typisierung auf und bezieht Informationen über das Verhalten anderer *kontextabhängig* und unter Berücksichtigung der Möglichkeit *unbeabsichtigten Betrugs* in den *probabilistischen Typglauben* ein.

Der Einsatz von *nicht-abstreitbaren Beweismitteln* führt zu einer Umgestaltung des *Empfehlungsmodells*, die die Möglichkeiten zu nicht wahrheitsgemäßen Empfehlungen beseitigt oder unvorteilhaft macht. Es werden zwei Arten von Beweismittel vorgestellt. Durch *transaktionale Beweismittel* dokumentieren sich Einheiten gegenseitig ihr Verhalten während einer Transaktion. Das Empfehlungssystem macht von diesen Beweismitteln Gebrauch um sicherzustellen, dass jede Einheit Betrugsverhalten meldet, sobald sie solches wahrnimmt. *Soziale Beweismittel* dienen dazu, beidseitige Vertrauensbeziehungen in Form von Bürgschaften sichtbar zu machen. Diese Bürgschaften werden in Selbstempfehlungen verbreitet. Der eigene Ansatz sorgt dafür, dass normative Einheiten durch die Angabe der eigenen Bürgen glaubwürdig ihren Typ signalisieren und sich somit von den strategischen Einheiten absetzen können.

Die *simulative Evaluation* des Ansatzes weist nach, dass die beiden Maximen des Entwurfs

eingehalten werden und damit Betrugsverhalten effektiv eingedämmt wird. Hierfür wird eine Methodik angewandt, die sich unter anderem darin auszeichnet, dass sie alle Arten von viel versprechendem Fehlverhalten mit Hilfe einer interaktiven Simulation aufspürt. Mehrere Sensibilitätsanalysen stellen heraus, welche Mindestanforderungen an die Systemumgebung gelten müssen, damit der Systementwurf erfolgreich ist. Für das Szenario universitärer Ad-hoc Netze zeigen die Simulationsergebnisse, dass die meisten Einheiten normativ am System teilnehmen und somit der Systementwurf erfolgreich ist. Abschließend werden weiterführende Konzepte vorgestellt, die den Ansatz dieser Arbeit für den Einsatz unter erschwerten Rahmenbedingungen erweitern.

Danksagung

Ich bin der Professorin Birgitta König-Ries dankbar, dass sie mich in ihr Forscherteam aufgenommen und meine wissenschaftlichen Arbeiten betreut hat. Sie war mir eine stets ansprechbare Betreuerin, die mir in den entscheidenden Momenten den Weg zur Promotion geebnet hat. Ich möchte ihr auch dafür danken, dass sie mir bei der Wahl meines Forschungsschwerpunkts und der Durchführung meiner Arbeiten freie Hand ließ.

Des Weiteren bin ich Professor Christof Weinhardt dankbar, dass er sich bereit erklärt hat, als zweiter Gutachter Zeit und Mühen in die Bewertung meiner Dissertation zu stecken.

Bei meinen Kollegen am IPD der Universität Karlsruhe möchte ich mich für ihr inhaltliches Feedback zu meiner Forschung bedanken. Besonders herausheben ist Michael Klein, mit dem ich dasselbe Büro geteilt habe. Er hat mich nicht nur in zahlreichen Diskussionen auf wichtige Ideen für meine Forschung gebracht. Darüber hinaus stand er mir bei Fragen des Layouts stets als kompetenter Ratgeber zur Verfügung.

Ich bin einigen weiteren Personen dankbar, die inhaltlich auf diese Dissertation eingewirkt haben: Tanja Nietschke und Professor Peter Reiher haben mir bei den Rechtsfragen, die in dieser Arbeit behandelt werden, entscheidende Hinweise gegeben. Die Gespräche mit dem Biologen Henri Saleh haben mir eine andere Sichtweise auf die Problemstellung meiner Arbeit aufgezeigt. Die Aneignung der wirtschaftswissenschaftlichen Vorkenntnisse, die dieser Arbeit zugrunde liegen, wurde von meinem Bruder Stephan angeregt. Schließlich sei auch all jenen Studenten gedankt, deren Arbeiten meine Forschung vorangetrieben haben. Stefan Fähnrich nimmt hierbei einen besonderen Platz ein, da er meine Forschung über den Zeitraum von zwei Jahren aktiv begleitet hat.

Erwähnt seinen an dieser Stelle auch all jene, die mich direkt oder indirekt dazu gebracht haben, mich zum Promotionsstudium zu entschließen: Guntram Gräf hat mich in meiner Zeit als Diplomand an das wissenschaftliche Arbeiten herangeführt. Nicht nur er sondern auch Professor Detlef Schmid hat mir den Schritt zum Promovieren nahe gelegt. In diesem Zusammenhang möchte ich auch Dr. Robert Wischhusen anführen, der mich während meiner Zeit als Gymnasialschüler gezielt gefördert hat.

Großen Einfluss auf meine Forschung hatte die Lektüre der chinesischen Klassiker von Konfuzius und Sun-tze's. Einige der Aussprüche dieser beiden Vordenker, die vor 2500 Jahren gelebt haben, sind als Zitate den Kapiteln dieser Arbeit vorgestellt. Insbesondere Konfuzius beschäftigte sich mit einem Problem, das dem meiner Arbeit sehr nahe kommt, nämlich dem Entwurf eines Normensystems für eine selbst-stabilisierende Gesellschaft.

Am Allermeisten möchte ich mich bei meiner Frau Sok Shee bedanken. Aus dem gemeinsamen Leben mit ihr habe ich die Kraft geschöpft, meine Promotion durchzuführen. Ich bin ihr sehr dankbar für ihr Vertrauen, ihre Ermunterungen, ihr Verständnis und vielem mehr. Auch meinen Eltern gilt besonderer Dank. Sie haben mich während meines Studiums fortwährend unterstützt und meinen Weg an die Hochschule überhaupt erst ermöglicht.

Inhaltsverzeichnis

Abbildungsverzeichnis

Tabellenverzeichnis

Teil I

Ausgangspunkt

Kapitel 1

人而無信 不知其可也

"Ich kann nicht erkennen, wie jemand etwas zu erreichen vermag, wenn sein Wort nicht vertrauenswürdig ist."
(Gespräche und Aussprüche des Konfuzius, 2.22)

Einführung

1.1 Motivation

In den letzten Jahrzehnten hat die Informationstechnologie eine immer wichtigere Rolle für die Gesellschaft und das Geschäftsleben eingenommen. Der ursprüngliche Zweck eines Informationssystems bestand darin, den Menschen beim Ablegen und Aufbereiten von Informationen zu unterstützen. Dazu griff jeder menschliche Benutzer auf sein eigenes Informationssystem zu, das auf einem Desktop-Computer bei ihm zu Hause oder an seiner Arbeitsstätte betrieben wurde. Im Universitätsumfeld stellte sich zum Beispiel die Situation für Studenten so dar, dass sie Informationen zu ihrem Universitätsstudium auf ihrem Computer ablegten und bearbeiteten. Hierunter fallen nicht nur Vorlesungs- und Übungsunterlagen sondern auch selbst erstellte Arbeiten und Dokumente. Informationen unterschiedlicher Benutzer wurden traditionell nur unter Einwirkung der Benutzer selbst ausgetauscht, typischerweise durch Übergabe von Speichermedien wie Disketten und später durch das Versenden von elektronischer Post. Dies bedeutete zum Beispiel für Studenten, dass sie selbst entscheiden mussten, von wem sie welche Informationen anfragen und welche Informationen sie anderen zur Verfügung stellen.

In letzter Zeit hat sich hier ein Wandel vollzogen. Wenn Menschen nicht mehr in den Informationsaustausch einbezogen werden, ergibt sich für sie eine Zeit- und Kostenersparnis. Um diese *Automatisierung* zu erreichen, muss ein menschlicher Benutzer im Informationssystem durch einen Computeragenten vertreten werden, der in der Lage ist, Entscheidungen im Sinne seines menschlichen Prinzipals zu fällen und umzusetzen. Die Automatisierung führt dazu, dass die menschlichen Benutzer und ihre Handlungsmöglichkeiten im Informationssystem nachgebildet werden. Für die *Kooperation* zwischen menschlichen Benutzern ist es nicht mehr zwingend, dass sie sich in der Realwelt kennen oder einen Kontakt aufbauen. Da ihre Computeragenten in ihrem Interesse im Informationssystem handeln, ist für menschliche Benutzer also nur noch die erhaltene Information als das Ergebnis der Kooperation ersichtlich und nicht, von wem sie auf welcher Weise beschafft wurde. Im Informationssystem entsteht damit eine *künstliche Gesellschaft*, deren *Einheiten* die Computeragenten sind.

In der Tradition des Client-Server-Paradigmas wird der Ansatz der Automatisierung dadurch realisiert, dass an einer *zentralen* Stelle auf dedizierter Hardware eine Software-Plattform instal-

liert wird, die die Computeragenten aufnehmen kann. Die menschlichen Benutzer müssen sich deswegen entfernt mit ihren Computeragenten koordinieren. Diese Architektur ist typisch für viele der heutigen Multi-Agenten Systeme [Woo02]. Sie hat zum Vorteil, dass genau kontrolliert werden kann, welche Computeragenten teilnehmen und ob sie sich gemäß der Regeln des Informationssystems verhalten. Jedoch wird dieser Vorteil durch *hohe Kosten* für die *Inbetriebnahme* und den *laufenden Betrieb* des Informationssystems erkauft. Dies liegt daran, dass der Betreiber des Informationssystems für die Beschaffung und den Betrieb der dedizierten Hardware und der Software-Plattform aufkommen muss. Zudem müssen die Benutzer dem Betreiber des Informationssystems vertrauen, dass er ihre Agenten in ihrem Sinne betreibt. Seine Vertrauenswürdigkeit kann der Betreiber jedoch nur durch eine Öffnung seines Systems gegenüber Kontrollen durch das Gemeinwesen (wie zum Beispiel eine staatliche Organisation) unter Beweis stellen. Damit kommen auf den Betreiber und die Benutzer weitere versteckte Kosten zu. Ein weiterer Nachteil des zentralen Ansatzes ist die entfernte Abstimmung zwischen menschlichen Benutzern und ihrer Computeragenten, die zeit- und kostenintensiv ist.

Eine Alternative zu zentralen Informationssystemen wird durch einen anderen Trend ermöglicht. Durch die Verbreitung von PCs und neuerdings auch Informationsgeräten (engl.: information appliances [Nor98]) wie Personal Digital Assistants (PDAs) liegt es für menschliche Benutzer nahe, mit ihrem eigenen Gerät am Informationssystem teilzunehmen. Dies gilt insbesondere deswegen, weil die Geräte über genügend Ressourcen verfügen, um nicht nur als reines Darstellungsgerät (engl.: thin client) einsetzbar zu sein. Außerdem ist die Kommunikationstechnik durch die Entwicklung des Internets und der drahtlosen Kommunikation (WLAN, Bluetooth) so weit fortgeschritten, dass die Geräte der menschlichen Benutzer in Peer-to-Peer (P2P) oder Ad-hoc Netzen direkt miteinander kommunizieren können. Im Vergleich zu dieser Kommunikation untereinander ist vor allem in Ad-hoc Netzen die Kommunikation mit einer zentralen Komponente erheblich aufwändiger. Insgesamt ist es also wünschenswert, dass jeder Computeragent auf dem Gerät seines jeweiligen menschlichen Prinzipals platziert wird. Damit entfällt nicht nur die entfernte Abstimmung zwischen menschlichen Benutzern und ihren Computeragenten. Darüber hinaus wird die zentrale Plattform für die Computeragenten verzichtbar. Die Kosten für die Inbetriebnahme und den laufenden Betrieb der zentralen Plattform entfallen also. Für den Einsatz des Informationssystems genügt es, an die menschlichen Benutzer eine Software zu verteilen, die sie auf ihrem eigenen Gerät installieren und die den Betrieb eines lokalen Computeragenten ermöglicht. Das Informationssystem wird *selbstorganisierend*, da die Geräte ohne Koordination einer zentralen Stelle kooperieren. Im Gegensatz zum zentralen Ansatz ist damit das Informationssystem *fehlertolerant* und besitzt keinen Flaschenhals für seine *Leistungsfähigkeit*. Die benötigten Ressourcen werden von den menschlichen Benutzern selbst durch den Einsatz ihrer Geräte zur Verfügung gestellt.

Es gibt aber ein Hindernis, diese Vision umzusetzen. Die Geräte der menschliche Benutzer und damit auch die Computeragenten entziehen sich jeder zentralen Kontrolle. Das Informationssystem ist also eine künstliche Gesellschaft, deren teilnehmenden Einheiten, die Computeragenten, *autonom* sind. Einem menschlichen Benutzer ist es nämlich möglich, seinen Computeragenten gezielt in seinem Sinne zu manipulieren. Dazu ist kein Expertenwissen erforderlich, da der Benutzer über das Internet an eine manipulierte Version der Systemsoftware gelangt sein kann [BH03]. Damit kann nicht mehr vorausgesetzt werden, dass sich die Computeragenten an die Regeln der Kooperation halten, die im Systementwurf vorgesehen sind. Vielmehr ist zu erwarten, dass Computeragenten sich gegenseitig betrügen. Ihre jeweiligen menschlichen Prinzipale müssen nämlich nicht derselben Organisation angehören und verfolgen daher Interessen, die untereinander in Konflikt stehen. Bei einem beidseitigen Austausch von Informationen kommt es zum Beispiel dadurch

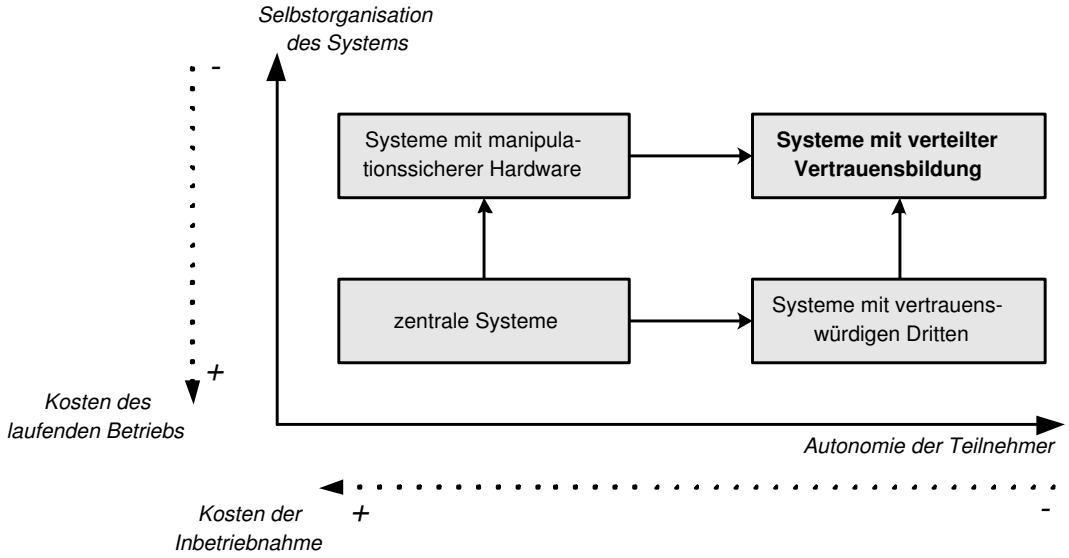

Abbildung 1.1: Erweiterung von Informationssystemen um Autonomie und Selbstorganisation

zum Betrug, dass einer der beiden Computeragenten dem anderen die versprochenen Informationen vorenthält. Ein betrügender Computeragent schont also seine Ressourcen und schneidet daher besser ab als sich kooperativ verhaltende Computeragenten. Letztendlich ist das Informationssystem dadurch zum Scheitern verurteilt, da es zu einem Zitronenmarkt [Ake70] degeneriert.

Um dieses Problem zu umgehen, werden üblicherweise zwei alternative Ansätze verfolgt. Sie schränken den Grad der Selbstorganisation oder der Autonomie ein, um Betrug durch die Computeragenten vorzubeugen. Diese Ansätze werden in Abbildung 1.1 eingeordnet. Der erste Ansatz besteht darin, die Autonomie der Computeragenten zu beseitigen. Dazu muss verhindert werden, dass menschliche Benutzer ihre Computeragenten manipulieren können. Es liegt also nahe zu fordern, dass Computeragenten auf *manipulationssicherer Hardware* [AK96] platziert werden. Dadurch lässt sich garantieren, dass sich die Computeragenten an die Regeln der Kooperation halten. Das Problem dieses Ansatzes liegt darin, dass die Geräte der menschlichen Benutzer zunächst einmal keine Manipulationssicherheit gewähren. Damit wird für die Teilnahme am Informationssystem der Erwerb zusätzlicher Hardware notwendig [BH03]. Folglich sind die Kosten für die Inbetriebnahme des Informationssystems hoch. Der zweite Ansatz verzichtet daher darauf, die Autonomie der Teilnehmer einzuschränken. Stattdessen wird die Selbstorganisation des Informationssystems aufgegeben. Der Ansatz besteht darin, dass eine zentrale Komponente die Einhaltung der Regeln der Kooperation überwacht. Sie fungiert als *Dritter* [Aso98], der im Falle von Betrug zwischen Computeragenten angerufen wird und den Betrug rückgängig machen oder zumindest nachhaltig bestrafen kann. Im Vergleich zum zentralen Informationssystem sind für diesen Ansatz die Kosten für die Inbetriebnahme des Informationssystems nicht so hoch. Jedoch wird weiterhin eine zentrale Komponente für den laufenden Betrieb benötigt. Um Betrug effektiv zu vermeiden, muss die zentrale Komponente ständig in Verbindung mit den Informationsgeräten der menschlichen Benutzer stehen. Eine solche fortwährende entfernte Abstimmung ist jedoch nur dann möglich, wenn die kostenpflichtigen Dienste externer Anbieter (wie zum Beispiel von UMTS [UMT03]) in Anspruch genommen werden. Weitere Kosten rühren daher, dass wie beim zentralen Ansatz der Betreiber für die Ausfallsicherheit und Leistungsfähigkeit der zentralen Komponente sorgen muss. Außerdem sind die Teilnehmer des Informationssystems von

der Vertrauenswürdigkeit des Betreibers durch entsprechende Kontrollmaßnahmen zu überzeugen. Die Kosten für den Betrieb des Informationssystems mit Dritten fallen daher nicht niedriger aus als die für den zentralen Ansatz. Keiner der beiden Ansätze schafft es also, die Vision eines selbstorganisierenden Informationssystems mit autonomen Teilnehmern umzusetzen.

Es ist also ein weiterer Ansatz nötig, der keine Kompromisse hinsichtlich des Grads der Selbstorganisation und der Autonomie verlangt. Wenn das Verhalten der Computeragenten weder von der Manipulationssicherheit noch von Dritten kontrolliert wird, müssen die teilnehmenden Computeragenten selbst diese Kontrolle ausüben (siehe Abbildung 1.1). Diese *soziale Kontrolle* [CCP98] besteht darin, dass Computeragenten durch Beobachtung vergangenen Verhaltens untereinander Vertrauen aufbauen. Die *Vertrauensbildung* ist im Interesse der Computeragenten selber, da sie dadurch die Wahrscheinlichkeit, betrogen zu werden, abschätzen können. Vertrauen ist also ein Ersatz für vollständige Information über das zukünftige Verhalten von potentiellen Kooperationspartnern [SY01]. Die Vertrauensbildung erfolgt *verteilt*, da sie von jedem Computeragenten ohne Hilfe eines Dritten durchgeführt wird. Sie beeinflusst maßgeblich Kooperationsentscheidungen: Ein Computeragent kooperiert nur dann mit anderen Computeragenten, wenn sie ihm vertrauenswürdig genug erscheinen. Dadurch, dass damit weniger Computeragenten zur Kooperation mit einem Betrüger bereit sind, entstehen Betrugskosten (engl.: defection costs). Der Ansatz der verteilten Vertrauensbildung gibt also einen *Anreiz*, im Zuge der Kooperation nicht zu betrügen. Der Ansatz kann zwar Betrug nicht verhindern, zielt aber darauf, dass Computeragenten es aus eigenem Interesse vermeiden zu betrügen. Damit könnte das Informationssystem auf die Geräte der menschlichen Benutzer verteilt werden, ohne dass zentrale Dritte oder manipulationssichere Hardware nötig würden.

Im Mittelpunkt dieser Arbeit steht also die Frage, ob und wie der Ansatz der verteilten Vertrauensbildung die Vision von selbstorganisierenden und auf autonomen Geräten verteilten Informationssystemen Realität werden lässt. Konkret wird ein System zur lokalen und empfehlungsgestützten Vertrauensbildung vorgeschlagen, dessen simulative Evaluation die Effektivität der Kooperationsanreize der verteilten Vertrauensbildung zeigt.

1.2 Szenario

Die Vision von Informationssystemen, die vollständig auf die Geräte der Endbenutzer verteilt sind, ist bisher noch recht abstrakt geblieben. In diesem Abschnitt wird die Bedeutung der Vision am Beispiel eines universitären Campus aufgezeigt. Die Vorstellung des *Campus-Szenarios* klärt nicht nur, warum die Teilnahme am Informationssystem für menschliche Benutzer von Vorteil ist, darüber hinaus wird auch untersucht, wie es zu Betrugsverhalten kommt und welche Schäden es für das Informationssystem mit sich bringt.

Die Eigenschaften des Campus-Szenarios werden anschließend in einem *Systemmodell* festgehalten. Dieses Systemmodell liegt den weiteren Teilen dieser Arbeit zugrunde.

1.2.1 Das Campus-Szenario

Die Vorstellung des *Campus-Szenarios* erfolgt entlang von vier Fragen:

1. Wie nehmen menschliche Benutzer (also Studenten oder Universitätsangestellte) an dem Informationssystem teil und warum wollen sie dies tun?

2. Wie sieht ein typischer Ablauf des Szenarios aus?

Abbildung 1.2: Das Campus-Szenario

3. Wie kommt es zu Betrugsverhalten im Informationssystem?

4. Warum wird das Bestehen des Gesamtsystems durch Betrugsverhalten gefährdet?

Im Folgenden werden diese Fragen nacheinander angegangen.

Teilnahme am Informationssystem. Die Universität sucht nach einer Möglichkeit, Studenten und Universitätsangestellte (kurz: Benutzer) beim Lernen und Arbeiten zu unterstützen. Dazu wird ein Informationssystem benötigt, durch das die Benutzer benötigte Informationsdienste erbracht bekommen. Aus Kostengründen soll dabei das Informationssystem von den Informationsgeräten (Laptops, PDAs) der Benutzer betrieben werden. Die Rolle der Universität beschränkt sich also darauf, ein Stück Software bereitzustellen, das sich interessierte Benutzer auf ihrem Informationsgerät installieren können. Während des Betriebs des Informationssystems erbringen die Informationsgeräte der teilnehmenden Benutzer gegenseitig die benötigten Informationsdienste, indem sie direkt miteinander in einem Ad-hoc-Netz kommunizieren. Technologisch wird dies dadurch ermöglicht, dass jedes Gerät eine WLAN-Karte gemäß dem Standard IEEE 802.11 [CWJS97] eingebaut hat und sie im Ad-hoc Modus verwendet. Da für den Betrieb des Informationssystems keine Basisstation benötigt wird, kann es sich auch auf Gebiete außerhalb des Campus-Geländes (zum Beispiel die Innenstadt, den benachbarten Park und die Wohngebiete der Stadt) erstrecken. Das Informationssystem bildet sich also immer an den Orten spontan heraus, wo Informationsgeräte der Benutzer sind. Dies wird in Abbildung 1.2 gezeigt.

Um den Nutzen der Teilnahme am Informationssystem des Campus-Szenarios zu veranschaulichen, werden im Folgenden drei Benutzer des Informationssystems vorgestellt:

- *Anna* besitzt einen PDA, den sie immer in ihre Vorlesungen mitnimmt. Darauf erstellt sie Vorlesungsmitschriebe und photographiert gelegentlich mit der im PDA eingebauten Kamera die Tafel des Unterrichtsraumes. Anna stellt ihren PDA aus, sobald sie nicht mehr mit ihm arbeitet, da die Batterie eventuell nicht für den Tag an der Universität ausreicht und sie ihn auch zu Hause aus Bequemlichkeit nicht gerne auflädt. Anna macht es nichts aus, dass andere ihre Vorlesungsmitschriebe erhalten, solange sie davon auch einen Vorteil hat. Sie benötigt den Mensaplan des Tages, da sie gerne schon in der letzten Vormittagsvorlesung darüber nachdenkt, welches Gericht sie in der Mensa nehmen wird. Außerdem braucht sie des Öfteren eine PDA-freundliche Darstellung der Vertiefungslektüre, die von den Dozenten empfohlen wurde. Wegen der beschränkten Ressourcen ihres PDAs kann sie aber ein entsprechendes Konvertierungsprogramm nicht auf ihrem Gerät installieren.

- *Bob* ist im Besitz eines Laptops, auf dem er in seiner Freizeit eine Reihe von Programmen installiert und elektronische Handbücher abgelegt hat. Bob geht selten in die Universität. Er benötigt daher nach Möglichkeit Vorlesungsunterlagen und -mitschriebe anderer. Außerdem kennt er sich mit den Fahrplänen der öffentlichen Verkehrsmittel nicht so gut aus, da er erst vor kurzem seinen Studienort gewechselt hat.

- *Claude* besucht im Laufe eines Tages Vorlesungen an verschiedenen Bereichen des Campus-Geländes. Da er seinen PDA regelmäßig auflädt, macht es Claude nichts aus, beim Durchqueren des Geländes seinen PDA anzulassen. Wie Bob ist auch Claude an Vorlesungsunterlagen und -mitschrieben interessiert. Außerdem ist er immer an den aktuellen Sonderangeboten der Geschäfte der Innenstadt interessiert.

Warum möchten Anna, Bob und Claude ihre Geräte an einem Informationssystem teilnehmen lassen, in dem Informationen ausgetauscht und Dienste füreinander erbracht werden? Die obige Auflistung der Wünsche und Fähigkeiten lässt erkennen, dass durch die Teilnahme für diese drei Benutzer ein Mehrwert entsteht. Dieser Mehrwert besteht auch nach längerem Betrieb des Informationssystems noch weiterhin. Wir zeigen dies anhand der Eigenheiten der drei Modellbenutzer:

- *Anna* erstellt wöchentlich Vorlesungsmitschriebe. Ihr Gerät ist daher eine Quelle von Informationen, deren Menge sich ständig erweitert. Selbst wenn Bob alle bisherigen Vorlesungsmitschriebe erhalten hat, ist er dennoch an denen der zukünftigen Vorlesungen interessiert.

- *Bobs* Gerät stellt Informationsdienste zur Verfügung, deren Erbringung auch nach mehrmaligem Aufruf nützlich für andere ist. Zum Beispiel wird Bobs Konvertierungsprogramm zur PDA-freundlichen Darstellung von Dokumenten von Anna immer dann benötigt, wenn sie die Vertiefungslektüre auf ihrem PDA lesen möchte. Sobald Anna eine neue Vertiefungslektüre erhält, ist der Dienst der Konvertierung für sie nützlich.

- *Claudes* Gerät nimmt an verschiedenen Orten am Informationssystem teil. Damit ermöglicht es den Austausch von Informationen zwischen zeitlich oder räumlich fernen Geräten. Zum Beispiel kann Bob in den Besitz von Annas Vorlesungsmitschrieben kommen, ohne dass Annas und Bobs Geräte je die Möglichkeit hätten, miteinander zu kommunizieren.

Zusammenfassend führt also die Teilnahme am Informationssystem deswegen zu einem Mehrwert für die Benutzer, weil Informationen und Ressourcen ungleich verteilt sind und die Geräte zeitlich/räumlich entfernt betrieben werden.

Typischer Ablauf. Wie sieht ein Zusammenspiel zwischen den Benutzern und zwischen ihren Geräten aus, bei dem dieser Mehrwert für die Teilnehmer geschaffen wird? Im Folgenden wird ein möglicher Tagesablauf beschrieben, bei dem die schon vorgestellten Studenten Anna, Bob und Claude die Hauptprotagonisten sind. Der Ablauf ist größtenteils aus der Sicht der menschlichen Benutzer beschrieben. In Abschnitt 2.1 werden wir kurz darauf eingehen, welche Technologien einsetzbar sind, um den beschriebenen Grad der Automatisierung zu erreichen.

Am Morgen geht *Anna* zu ihrer Informatik-Vorlesung. Sie hat ihren PDA mitgenommen, mit dessen Hilfe sie die Vorlesung kommentiert und ab und zu Schnappschüsse des Tafelanschriebs erstellt. Während dieser Zeit ist *Claude* im Gebäude der Wirtschaftswissenschaften. Er schaut sich auf seinem PDA die Sonderangebote an, die sein PDA am Morgen beim Durchqueren der Innenstadt aufgesammelt hat. Claudes PDA ist nämlich von Claude so eingestellt worden, dass er Vorlesungsmitschriebe und Sonderangebote sammelt. Kurz darauf macht sich Claude in Richtung Informatik-Bau auf, da er dort bald eine Vorlesung besuchen muss. Auf dem Weg dorthin gerät er in die Nähe[1] eines anderen Studenten *David*, der sich zuvor in der Mensa aufgehalten hat. Davids Gerät hat dort den Mensaplan des Tages erhalten. Da sowohl Claude als auch David ihre Geräte nicht ausgeschaltet haben, bemerken sich ihre Geräte gegenseitig. Die Geräte erkennen zudem die Möglichkeit eines beidseitig vorteilhaften Informationsaustauschs. Sie initiieren eine *Transaktion*: Die *Aktion* von Claudes Gerät besteht darin, die gesammelten Sonderangebote zu übermitteln, an denen auch David interessiert ist. Davids Gerät übergibt als Gegenleistung den Mensaplan. Die Studenten Claude und David erfahren nichts von dem Tausch. Jedoch werden sie von ihren Geräten jeweils über die Verfügbarkeit neuer gewünschter Information benachrichtigt.

Währenddessen nähert sich Annas Vorlesung dem Ende. Anna überlegt sich, was sie am Mittag in der Mensa für Wahlmöglichkeiten bekommt. Sie beauftragt daher ihren PDA, eine solche Information einzuholen. Zu diesem Zeitpunkt erreicht Claude den Informatik-Bau. Annas und Claudes PDAs bemerken sich gegenseitig und erkennen die Möglichkeit einer Transaktion: Claudes Mensaplan wird gegen Annas neuen Vorlesungsmitschrieb getauscht.

Nach der Mensa steigt Anna in die Straßenbahn ein, um nach Hause zu fahren. In der Straßenbahn versucht sie, auf ihrem PDA die Vertiefungslektüre durchzulesen, die am Morgen vom Dozenten verteilt wurde. Jedoch gestaltet sich die Lektüre als schwierig, da das Dokument nicht für die Darstellung auf dem PDA ausgelegt ist. Anna beauftragt daher ihr Gerät, nach Möglichkeit eine PDA-freundliche Darstellung der Vertiefungslektüre zu besorgen. Zu dieser Zeit verlässt *Bob* sein Zuhause und begibt sich zur nächsten Straßenbahnhaltestelle. Da Bob die morgendliche Vorlesung verpasst hat, beauftragt er seinen Laptop, Vorlesungsunterlagen und -mitschriebe Anderer zu sammeln. An der Haltestelle bemerken sich Annas und Bobs Geräte gegenseitig und führen eine Transaktion durch: Bobs Laptop ist in der Lage, Annas Vertiefungslektüre in eine PDA-freundliche Fassung zu konvertieren. Im Gegenzug besteht die Aktion von Annas PDA daraus, den Vorlesungsmitschrieb Bob zur Verfügung zu stellen.

Bob begibt sich anschließend in den Schlosspark, um dort auf der Wiese mit seinem Laptop zu arbeiten. In seiner Nähe befindet sich ein weiterer Student, *Ernst*, der auf seinem Laptop in Java programmiert. Als Bob am Abend kurz davor ist zu gehen, fragt er sich, wann seine Straßenbahn abfahren wird. Er beauftragt sein Gerät diese Information einzuholen. Zu dieser Zeit stößt Ernst auf ein Problem mit der Serialisierung von Java-Objekten und beauftragt daher sein Gerät, den entsprechenden Teil des Java Referenzhandbuchs zu besorgen. Da Bobs Laptop das

[1] Die zwei Studenten müssen nicht in unmittelbarer Nähe oder in Sichtweite sein. Mit gängigen WLAN Technologien können Geräte über eine Entfernung von circa 50 Meter kommunizieren [JKSS04]. Bei dem neueren Standard IEEE 802.11g gilt dies auch bei Kommunikation über Hindernisse wie Mauern hinweg [ZEE03]. Ohne Hindernisse sind dann sogar 100 Meter möglich.

Referenzhandbuch und Ernsts Laptop den Straßenbahn-Fahrplan gespeichert haben, bemerken die Geräte die Möglichkeit einer beidseitig vorteilhaften Transaktion und führen sie durch. Bob bekommt von seinem Laptop die Fahrzeiten angezeigt und Ernst kann im Referenzhandbuch nachlesen.

Herkunft von Betrugsverhalten. Die Beschreibung eines typischen Ablaufes im Informationssystem ging davon aus, dass die Studenten die *originale Systemsoftware*, wie von der Universität herausgegebenen, auf ihre Geräte installiert haben. Wenn Studenten hingegen diese Systemsoftware manipulieren, ändert sich die Situation im Informationssystem. Zur Veranschaulichung betrachten wir im Folgenden den Studenten *Manuel*.

Manuel möchte auch mit seinem Gerät am Informationssystem teilnehmen. Allerdings gibt er sich mit dem zu erwartenden Mehrwert der Kooperation nicht zufrieden. Vor der Installation der *originalen Systemsoftware*, ändert er sie in seinem Sinne. Dazu dekompiliert er sie und ändert das Programm an entscheidenden Stellen. Nach der Installation dieser *manipulierten Version* der Systemsoftware unterscheidet sich das Verhalten von Manuels Gerät in zweierlei Hinsicht von dem der anderen Geräte. Einerseits gibt sein Gerät vor, alle von anderen benötigten Informationen zu besitzen. Andererseits betrügt sein Gerät Transaktionspartner, indem es nach Erhalt der erwünschten Informationen darauf verzichtet, die versprochene Gegenleistung zu erbringen. Die Auswirkung dieser Manipulation verdeutlichen wir beispielhaft an einem Punkt des beschriebenen Szenario-Ablaufs: Wenn Manuel sich am Morgen auch im Informatik-Bau befindet und an Annas Vorlesungsmitschrieb interessiert ist, gibt sein Gerät vor, den von Anna gewünschten Mensaplan zu haben. Annas und Manuels Geräte initiieren daher eine Transaktion. Nachdem Manuels Gerät den Vorlesungsmitschrieb erhalten hat, bricht es die Verbindung ab. Damit betrügt Manuels Gerät Annas Gerät um den versprochenen Mensaplan.

Warum manipuliert Manuel sein Gerät, so dass es andere Geräte betrügt? Das liegt daran, dass dann sein Gerät weitaus weniger Ressourcen benötigt und verbraucht. Im Vergleich zu Anna und Bob muss Manuel keine Informationen oder Informationsdienste bereitstellen. Außerdem hat Manuel auch einen Vorteil gegenüber Claude: Im Szenario-Ablauf tauscht Claudes Gerät den Mensaplan ein und speichert ihn, um durch seine Übermittlung an Annas Gerät den Vorlesungsmitschrieb zu erhalten. Im Gegensatz dazu kommt Manuels Gerät an den Vorlesungsmitschrieb, ohne sich um den Mensaplan kümmern zu müssen.

Als Trittbrettfahrer ist Manuel also im Informationssystem sehr erfolgreich. Es liegt daher nahe, dass andere Studenten Manuel nachahmen. Dazu sind nicht unbedingt technische Kenntnisse zur Manipulation der Systemsoftware notwendig. Studenten, die zur Installation einer manipulierten Version der Systemsoftware bereit sind, können diese auch von Manuel selbst erhalten. Als Folge benutzen mehr und mehr am System teilnehmende Geräte manipulierte Software.

Folgen von Betrugsverhalten. Durch die Vorteilhaftigkeit von Betrugsverhalten müssen wir erwarten, dass viele der teilnehmenden Benutzer auf ihren Geräten manipulierte Versionen der Systemsoftware installiert haben. Dies hat zwei Auswirkungen auf das Informationssystem. Erstens trifft ein Gerät in einer Transaktion fast ausschließlich auf Geräte, die manipulierte Software benutzen. Dies liegt daran, dass diese im Gegensatz zu Geräten mit der originalen Systemsoftware vorgeben, über Informationen zu verfügen, die sie nicht haben. Wenn ein Gerät eine bestimmte Information sucht, wird es also sehr wahrscheinlich auf Geräte mit manipulierter Software aufmerksam. Zweitens wird ein Gerät im Laufe der Transaktion fast immer betrogen. Dies hängt mit dem ersten Punkt zusammen. Als Transaktionspartner werden wie gezeigt fast ausschließlich die Geräte mit manipulierter Software gewählt. Diese betrügen jedoch immer. Es ergibt sich

insgesamt also, dass Betrug durch den Transaktionspartner zu erwarten ist.

Wie wirkt sich diese Feststellung auf das Informationssystem aus? Im Folgenden beschreiben wir die Auswirkungen aus der Sicht der drei Hauptprotagonisten Anna, Bob und Claude:

- *Annas* PDA versucht wiederholt, den Mensaplan (beziehungsweise die PDA-freundliche Darstellung der Vertiefungslektüre) zu erhalten. Jedoch wird ihr Gerät nach dem Übermitteln des Vorlesungsmitschriebs von den Transaktionspartnern darum betrogen. Anna bemerkt, dass ihr PDA durch diese vergeblichen Versuche langsamer auf ihre Kommandos reagiert und sich die Batterie durch die wiederholte Übertragung der Vorlesungsunterlagen entleert. Auf der anderen Seite kann Anna keinen Mehrwert mehr erkennen, am Informationssystem teilzunehmen. Es bleibt ihr also nichts anderes übrig, als ihren PDA auszuschalten oder die Systemsoftware zu deaktivieren. Da diese Erfahrung sich im Laufe der Tage wiederholt, entscheidet sich Anna schließlich, die Systemsoftware ganz zu deinstallieren und nicht mehr am Informationssystem teilzunehmen.

- *Bobs* Laptop versucht vergeblich, durch Ausführung von Informationsdiensten wie dem Konvertieren von Dokumenten zu dem erwünschten Vorlesungsmitschrieb zu kommen. Selbst als Bob in die Nähe von Anna gerät, ergibt sich keine Transaktion, da Anna schon zuvor die Systemsoftware auf ihrem PDA deaktiviert hat. Für Bob ist wie auch zuvor für Anna die langsame Reaktion seines Geräts und die Entleerung seiner Batterie ärgerlich. Er zieht daher denselben Schluss wie Anna und deinstalliert letztendlich die Systemsoftware von seinem Laptop.

- *Claudes* PDA gibt auf dem Weg durch das Campus-Gelände an viele andere Geräte die ihm bekannten Sonderangebote weiter. Jedoch werden von den Gegenleistungen, die Claudes PDA von den Transaktionspartnern versprochen werden, keine erbracht. Claude bemerkt beim Erreichen des Informatik-Baus, dass die Batterie seines PDAs bereits leer ist. Abends sieht Claude zu Hause beim Wiederaufladen seines PDAs, dass er trotzdem nicht den erwünschten Vorlesungsmitschriebe oder neue Sonderangebote erhalten hat. Claude wird daher am folgenden Tag seinen PDA beim Erreichen des Campus-Geländes ausschalten und damit nicht mehr am Informationssystem teilnehmen.

Es ergibt sich also, dass alle Benutzer, die zum Informationssystem etwas beitragen, zum Verlassen des Informationssystems gebracht werden. Dieses Phänomen wird in der Literatur als *adverse Selektion* bezeichnet [BS89]. Das Informationssystem degeneriert zu einem *Zitronenmarkt* [Ake70], in dem nur noch Geräte teilnehmen, die manipulierte Versionen der Systemsoftware installiert haben. Selbst die Benutzer solcher Geräte (wie zum Beispiel Manuel) verlassen letztendlich das System, da keine interessanten Informationen oder Informationsdienste mehr im Umlauf sind. *Damit ist das Ziel gescheitert*, ein Informationssystem zur Unterstützung der universitären Benutzer einzusetzen.

Fazit. Das Informationssystem des Campus-Szenarios ist für menschliche Benutzer attraktiv, da es ihnen bei der Teilnahme einen Mehrwert bietet. Dieser wird dadurch geschaffen, dass das Informationssystem die ungleiche Verteilung von Informationen und Ressourcen und den zeitlich/räumlich entfernten Betrieb der Informationsgeräte überbrückt.

Durch die Manipulation der Systemsoftware können jedoch menschliche Benutzer ihre Geräte dazu bringen zu betrügen. Da dies einen Vorteil für den Benutzer darstellt, wird Betrugsverhalten im Informationssystem allgegenwärtig. Schließlich scheitert daran das Informationssystem.

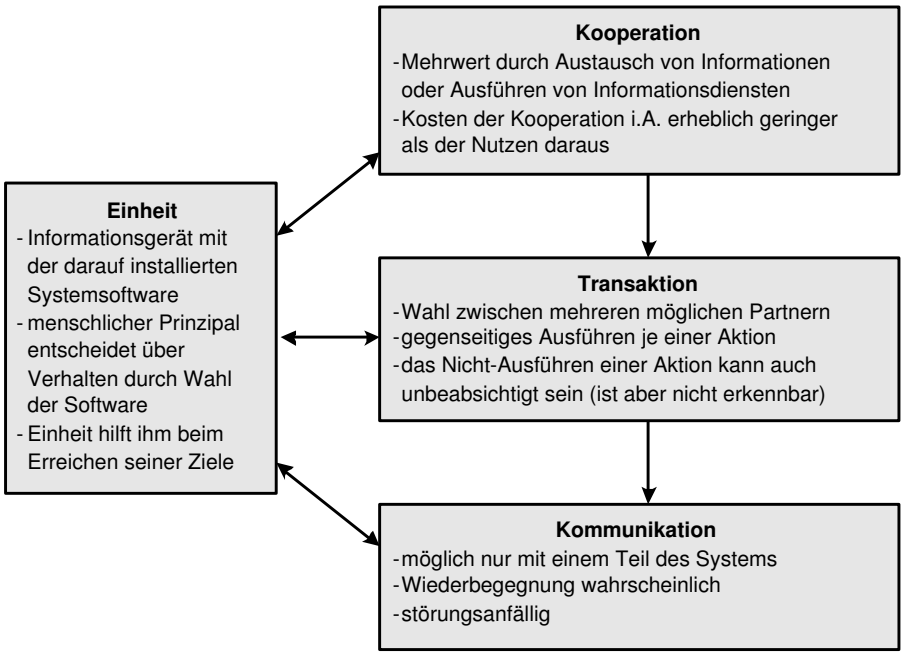

Abbildung 1.3: Schematische Darstellung des Systemmodells

Es wird also ein Ansatz benötigt, der die Manipulation der Systemsoftware und damit Betrugsverhalten im Informationssystem vermeidet. Nur dann kann das Informationssystem für die teilnehmenden Benutzer einen Mehrwert schaffen.

1.2.2 Abgeleitetes Systemmodell

Im Folgenden wird ein *Systemmodell* abgeleitet, das die Eigenschaften des Campus-Szenarios festhält. Mit dem Aufstellen des Systemmodells verfolgen wir zwei Ziele:

1. Die entscheidenden Rahmenbedingungen des Szenarios werden herausgestellt. Damit wird deutlich, welche Eigenschaften einen Ansatz zur Betrugsvermeidung erleichtern oder erschweren.

2. Da wir das Systemmodell den weiteren Teilen dieser Arbeit zugrunde legen, sind wir in der Lage, später eingeführte Mechanismen auf einer abstrakten Ebene präzise zu beschreiben.

Die Vorstellung des Systemmodells gliedert sich in vier Teile. Eine Übersicht dieser Teile und ihres Zusammenspiels wird in Abbildung 1.3 gegeben.

Einheiten. Um im Campus-Szenario am Informationssystem teilnehmen zu können, muss ein menschlicher Benutzer ein Informationsgerät besitzen, auf dem die Systemsoftware installiert ist. Ist diese Voraussetzung erfüllt, so kann das Gerät mit Hilfe der Software im Informationssystem handeln. Das Gerät mitsamt der installierten Systemsoftware stellt also eine *Einheit* des Informationssystems dar. Der *menschliche Prinzipal* einer Einheit ist der Benutzer, der durch sie am Informationssystem teilnimmt. Jede Einheit weiß von den Zielen des eigenen menschlichen Prinzipals (Anna muss im Szenario ihren PDA davon unterrichten, dass sie am Mensaplan interessiert

ist) und versucht im Informationssystem, diese Ziele zu verfolgen (der PDA besorgt sich den Mensaplan von Claudes Gerät).

Das Verhalten einer Einheit wird davon bestimmt, welche Version der Systemsoftware auf dem Gerät installiert ist. Die *originale Systemsoftware* wird vom *Systementwerfer* selbst (im Szenario die Universität) bereitgestellt. Jedoch kann es auch *manipulierte Versionen* der Systemsoftware geben (wie die von Manuel erstellte). Durch die Wahl, welche Software er auf sein Gerät installiert, entscheidet ein menschlicher Benutzer damit über das Verhalten seines Geräts. Diese Wahl kann nur vom Benutzer des Gerätes selbst und nicht etwa von anderen Benutzern getroffen werden (im Szenario kann niemand Manuel daran hindern, eine manipulierte Version der Software zu benutzen). Die Teilnehmer des Informationssystems sind also untereinander *autonom*.

Kooperation. Bei der Beschreibung des Campus-Szenarios haben wir bereits festgestellt, dass eine Information (Annas Vorlesungsmitschrieb) oder ein Informationsdienst (Bobs Konvertierungsprogramm) sich nur auf einem Teil der Einheiten befindet. Eine Einheit muss jedoch unter Umständen Informationen oder Informationsdienste in Anspruch nehmen, um die Ziele ihres menschlichen Prinzipals zu verfolgen. Der Austausch von Informationen oder Informationsdiensten stellt *Kooperation* zwischen den Einheiten dar. Dabei ist der *Nutzen* aus der Kooperation *wesentlich größer* als die *Kosten* dafür. Die Teilnahme am Informationssystem führt dadurch zu einem Mehrwert für die menschlichen Benutzer. Im Szenario bekommt zum Beispiel Claudes Gerät im Laufe der Kooperation mit anderen Geräten die Sonderangebote, an denen Claude interessiert ist, und Annas Vorlesungsmitschrieb. Dafür muss Claude lediglich sein Gerät während des Tages anlassen und es dadurch eventuell etwas früher wieder aufladen. Ähnlich verhält es sich mit Bob, der die etwas langsamere Reaktion seines Laptops während der Konvertierung von Annas Vertiefungslektüre in Kauf nimmt, da er dafür den gewünschten Vorlesungsmitschrieb erhält.

Kommunikation. Um untereinander zu kooperieren, müssen Einheiten letztlich in der Lage sein, Nachrichten auszutauschen. Diese *Kommunikation* ermöglicht es, dass Einheiten Gelegenheiten der Kooperation erkennen und ausnutzen.

Allerdings können Einheiten nicht beliebig miteinander kommunizieren. Aufgrund der beschränkten Sendereichweite der Geräte ist es für eine Einheit im Allgemeinen nur möglich, mit einem *kleinen Teil* der anderen Einheiten des Systems zu kommunizieren. So werden zum Beispiel im Szenario Claudes Gerät und Annas PDA erst dann aufeinander aufmerksam, als Claude den Informatik-Bau erreicht, in dem sich Anna befindet. Die Menge der erreichbaren Einheiten ändert sich ständig, da die menschlichen Benutzer sich mit ihren Geräten bewegen und die Geräte zu unterschiedlichen Zeitpunkten angeschaltet sind. Dadurch ist die Zahl der Einheiten, mit denen insgesamt kooperiert werden kann, weitaus größer als die Zahl derer, die sich zu einem gewissen Zeitpunkt in Kommunikationsreichweite befinden. So kann zum Beispiel Claudes Gerät von Daniel Sonderangebote und von Anna den Vorlesungsmitschrieb erhalten, obwohl Daniels und Annas Geräte nicht gleichzeitig von Claudes Gerät erreichbar sind.

Ein wichtiger weiterer Punkt ist, dass die *Wiederbegegnung* zweier Einheiten *wahrscheinlich* ist. Das liegt daran, dass die Bewegung und die Aktivitäten ihrer menschlichen Prinzipale gewissen zeitlichen und örtlichen Regelmäßigkeiten unterworfen ist. Im Szenario rühren diese Regelmäßigkeiten zum Beispiel vom Stundenplan oder von den Gewohnheiten der Studenten her.

Ein wesentliches Kennzeichen für die Kommunikation zwischen Einheiten ist, dass sie *störungsanfällig* ist. Es kann jederzeit vorkommen, dass eine Nachricht nicht den designierten Empfänger erreicht. Durch die Verwendung passender Transportprotokolle können zwar vorübergehende

Störungen[2] als die Ursache des Nachrichtenverlusts ausgeschlossen werden. Ein Nachrichtenverlust kann aber auch daher rühren, dass während der Kommunikation die beteiligten Einheiten außer Reichweite gelangen. Dies geschieht zum Beispiel im Szenario, wenn Claude den Informatik-Bau verlässt, bevor sein Gerät den Vorlesungsmitschrieb von Annas PDA erhalten hat. Die Wahrscheinlichkeit für einen solchen Nachrichtenverlust hängt von der Übertragungszeit einer Nachricht ab. Je größer sie ist, desto wahrscheinlicher ist es, dass einer der menschlichen Prinzipale während ihrer Übertragung außer Reichweite gelangt. Bei den meisten zu übertragenden Nachrichten ist jedoch die Wahrscheinlichkeit eines Nachrichtenverlusts sehr gering[3].

Transaktionen. Als elementarer Bestandteil der Kooperation übergibt eine Einheit eine Information an eine andere Einheit oder erbringt einen Informationsdienst für sie. Dies bezeichnen wir im Folgenden als das Ausführen einer *Aktion*. So führt zum Beispiel Annas PDA eine Aktion für Claudes Gerät aus, indem er Annas Vorlesungsmitschrieb übergibt. Bei der Betrachtung von Kooperation im Campus-Szenario wird klar, dass Kooperation immer als eine *Transaktion* zwischen zwei Einheiten stattfindet, die sich *gegenseitig* je eine Aktion ausführen. Eine Einheit hat eine *Transaktionsgelegenheit* mit einer weiteren Einheit (dem *potentiellen Transaktionspartner*), wenn beide Einheiten je eine Aktion benötigen, die jeweils der andere ausführen kann. Zum Beispiel hat Annas PDA eine Transaktionsgelegenheit mit Claudes Gerät, da Annas PDA und Claudes Gerät den Mensaplan beziehungsweise den Vorlesungsmitschrieb benötigen. Für eine Transaktionsgelegenheit kann eine Einheit auch *mehrere potentielle Transaktionspartner* haben. Im Szenario ist das zum Beispiel dann der Fall, wenn auch Daniel den Informatik-Bau erreicht und am Vorlesungsmitschrieb Annas interessiert ist. Dann hat Annas PDA die Wahl, ob er den Mensaplan von Claudes oder Daniels Gerät bezieht. Da sich Informationen schnell über den Campus verteilen können, ist es sogar ziemlich wahrscheinlich, dass es für eine Transaktionsgelegenheit mehrere potentielle Transaktionspartner gibt.

Das Ausführen einer Aktion kann einer Einheit *unbeabsichtigterweise* misslingen. Dies liegt an zweierlei Gründen: Erstens besteht vor allem bei der Erbringung von Informationsdiensten (die Konvertierung der Vertiefungslektüre durch Bobs Laptop) die Gefahr, dass durch Ressourcenmangel (zu wenig freier Hauptspeicher) oder Korrumpierung der Information (eine Datei mit unbekanntem Format ist unvollständig) das Ausführen der versprochenen Aktion fehlschlägt. Zweitens kann ein Nachrichtenverlust dazu führen, dass das Ausführen einer Aktion unbeabsichtigterweise misslingt. Dies liegt daran, dass das Übermitteln einer Information ein integraler Bestandteil einer Aktion ist. Zum Beispiel besteht die Aktion von Annas PDA darin, den Vorlesungsmitschrieb an Claudes Gerät zu übermitteln. Wenn durch einen Verbindungsabbruch die Nachricht, die den Mitschrieb beinhaltet, verloren geht, ist die Aktion des PDAs fehlgeschlagen. Wie wir bereits gesehen haben, ist die Wahrscheinlichkeit des Nachrichtenverlusts sehr gering. Entsprechendes gilt daher auch für das unbeabsichtigte Nichtausführen einer Aktion.

Zur Möglichkeit des unbeabsichtigten Fehlverhaltens kommt eine weitere Eigenschaft, die jedweden Ansatz zur Betrugsvermeidung erschwert: Eine Einheit, die unbeabsichtigterweise eine Aktion nicht ausführt, *weiß nichts* von diesem Misslingen. Wir verdeutlichen diesen Punkt an

[2] Bei den drahtlosen Technologien, die im Campus-Szenario zum Einsatz kommen, werden solche vorübergehende Störungen zum Beispiel von den Übertragungssignalen anderer Geräte oder von kleinen Hindernisse im Übertragungsweg verursacht.

[3] Da mit WLAN 802.11 Übertragungsraten von 100 KiloByte Nutzdaten pro Sekunde üblich sind [JKSS04], dauert zum Beispiel die Übertragung eines Mensaplans nur den Bruchteil einer Sekunde. Entsprechend gering ist daher die Wahrscheinlichkeit, dass gerade zu diesem Moment ein Verbindungsabbruch stattfindet. Die Übertragung von Annas Vertiefungslektüre (üblicherweise um die 100 KiloByte) dauert ungefähr eine Sekunde, so dass auch hier ein Nachrichtenverlust durch Verbindungsabbruch sehr unwahrscheinlich ist.

einem Beispiel aus dem Bereich des Campus-Szenarios: Nehmen wir an, dass Annas PDA im Begriff ist, an Claudes Gerät den Vorlesungsmitschrieb zu schicken. Als Annas PDA eine entsprechende Nachricht übermittelt, geraten beide Geräte außer Reichweite. Annas PDA kann also keine bestätigende Antwort von Claudes Gerät bekommen. Dennoch kann Annas PDA daraus nicht schließen, dass Claudes Gerät den Vorlesungsmitschrieb nicht erhalten hat. Wir zeigen das an einer parallelen Betrachtung: Nehmen wir an, dass Annas PDA nicht mit Claudes sondern mit Manuels Gerät in eine Transaktion geht. Weiterhin gehen wir davon aus, dass Annas PDA erfolgreich den Vorlesungsmitschrieb an Manuels Gerät übermittelt. Nun hat Manuel die Systemsoftware auf seinem Gerät so manipuliert, dass sie das Versenden jedweder bestätigenden Antwort unterlässt. Er tut dies, damit sein Gerät versprochene Gegenleistungen (im Beispiel der Mensaplan) nicht erbringen muss. Dadurch tarnt er das Betrugsverhalten seines Geräts. Wie zuvor bei der Transaktion mit Claudes Gerät erhält Annas PDA daher keine Bestätigung. Wir schließen also daraus, dass beim Ausbleiben einer bestätigenden Antwort drei Möglichkeiten existieren: **(1)** Der Transaktionspartner hat die Information nicht erhalten, obwohl sie ihm zugesendet worden ist. Gemäß dem Transaktionsmodell ist aber das erfolgreiche Zusenden ein integraler Bestandteil der eigenen Aktionsausführung. Wir sprechen in diesem Fall also davon, dass die Aktion unbeabsichtigterweise nicht ausgeführt worden ist. **(2)** Die bestätigende Antwort ist verloren gegangen. **(3)** Der Transaktionspartner hat die bestätigende Antwort unterschlagen. Die Schwierigkeit liegt also darin, dass zwischen dem Nachrichtenverlust aus (1) und (2) und dem Betrugsverhalten aus (3) nicht unterschieden werden kann.

1.3 These

In den vorigen Abschnitten haben wir die Vision von einem Informationssystem vorgestellt, das vollständig auf die Informationsgeräte der teilnehmenden menschlichen Benutzer verteilt ist. Das Campus-Szenario zeigt, dass solche Informationssysteme wünschenswert und sinnvoll sind. Allerdings zeigt das Szenario auch Folgendes: Ohne einen Ansatz zur *effektiven Betrugsvermeidung* kann ein solches Informationssystem nicht existieren und damit unsere Vision nicht realisiert werden.

Unsere Intuition, dass diese Vision dennoch realisiert werden kann, geben wir in der folgenden These zum Ausdruck:

These:
Durch einen passenden Ansatz zur Betrugsvermeidung lässt sich unsere Vision realisieren. Das bedeutet, dass das Informationssystem selbstorganisierend betrieben werden kann, indem es vollständig auf die Informationsgeräte der autonomen Teilnehmer verteilt wird, ohne dass seine Existenzfähigkeit gefährdet ist.

Die These fordert also außer der effektiven Betrugsvermeidung, dass wie im Campus-Szenario die Teilnehmer autonom sind und das Informationssystem selbstorganisierend ist.

Wie in der Motivation kurz angedeutet (und in den kommenden Abschnitten 2.2 und 2.3 genau besprochen), sind Ansätze mit manipulationssicherer Hardware und mit Dritten in der Lage, Betrugsverhalten effektiv zu vermeiden. Jedoch schränken sie dafür die Autonomie der Teilnehmer und die Selbstorganisation des Informationssystems ein. Diesen Nachteil haben Ansätze der verteilten Vertrauensbildung nicht. Bei ihnen ist es aber fraglich, ob sie Betrugsverhalten effektiv vermeiden können. Es scheint also so, als ob es einen *Zielkonflikt* zwischen effektiver Betrugsvermeidung einerseits und den Systemeigenschaften der Autonomie und Selbstorganisation andererseits gibt. Dieser Zielkonflikt wird in Abbildung 1.4 dargestellt.

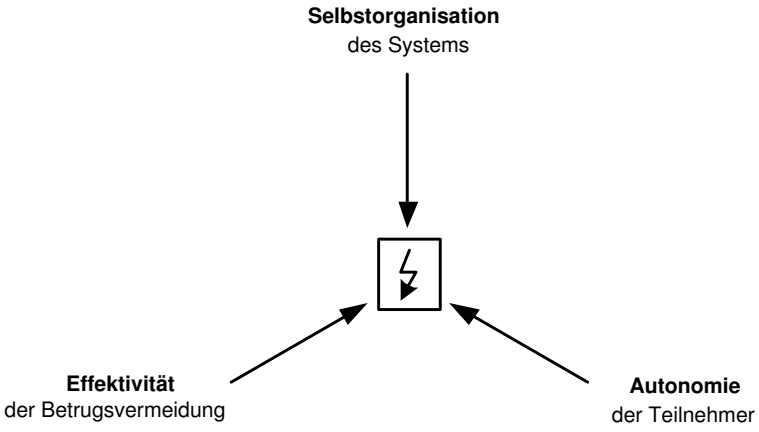

Abbildung 1.4: Zielkonflikt für die Realisierung der Vision

Wann sind wir in der Lage, unsere These zu validieren? Die effektive Betrugsvermeidung ist nur ein notwendiges Kriterium dafür, dass das Informationssystem existenzfähig ist. Damit ist nämlich noch nicht geklärt, ob menschliche Benutzer trotz der Einschränkung der Autonomie und Selbstorganisation (und den damit verbundenen Kosten) bereit sind, am Informationssystem teilzunehmen. Zur Validierung der These müssen wir uns also der Herausforderung stellen, den *Zielkonflikt* zwischen Selbstorganisation des Informationssystems, Autonomie der teilnehmenden Einheiten und Effektivität der Betrugsvermeidung zu lösen.

1.4 Lösungsansatz

Der Ansatz dieser Arbeit zielt auf die Validierung der These, die im letzten Abschnitt vorgestellt worden ist. Dazu muss der Zielkonflikt zwischen Selbstorganisation des Informationssystems, Autonomie der teilnehmenden Einheiten und Effektivität der Betrugsvermeidung gelöst werden.

Im Kern sorgt der Ansatz dafür, dass sich die Einheiten des Informationssystems aus eigenem Antrieb gegenseitig kontrollieren. Die dadurch entstehende *verteilte Vertrauensbildung* dämmt auf effektive Weise Betrugsverhalten ein, ohne die geforderte Selbstorganisation des Informationssystems und die Autonomie der Teilnehmer zu kompromitieren. Ein eigener Ansatz ist insofern notwendig, als die existierenden Vorschläge zur verteilten Vertrauensbildung nicht ausgereift sind.

Der Entwurf der verteilten Vertrauensbildung ist nur dann erfolgreich, wenn die Geräte von potentiell interessierten menschlichen Benutzern mit der Software des Systementwerfers am System teilnehmen. Hieraus ergeben sich zwei Maximen: Einerseits muss der Entwurf hinreichend gutartiges Verhalten in Normen vorschreiben. Andererseits dürfen Einheiten keinen Anreiz haben, die Software zu manipulieren, um sich opportunistisch zu verhalten. Dies führt zur Typisierung der Einheiten in normative und strategische Einheiten, die die ursprüngliche beziehungsweise manipulierte Software verwenden.

In dieser Arbeit wird ein Modell aufgestellt, wie jede Einheit basierend auf ihren eigenen Transaktionserfahrungen ihren Glauben über den Typ anderer Einheiten aufbauen und damit ihre Vertrauensentscheidungen fällen kann. Dieses Modell ist probabilistisch fundiert und berücksichtigt die Möglichkeit unbeabsichtigten Fehlverhaltens. Um auch von den Transaktionserfahrungen Anderer profitieren zu können, wird das Modell der Glaubensbildung um ein Empfehlungssystem erweitert. Dieses überführt das Konzept der Nichtabstreitbarkeit in den Kontext der verteilten

Vertrauensbildung. Damit werden die Möglichkeiten zu nicht wahrheitsgemäßem Empfehlungen beseitigt oder unvorteilhaft gemacht. Es werden zwei Arten von nicht-abstreitbaren Marken, die Beweismittel genannt werden, eingeführt und in den eigenen Ansatz der verteilten Vertrauensbildung integriert: **(1)** Durch transaktionale Beweismittel dokumentieren sich Einheiten gegenseitig ihr Verhalten während einer Transaktion. Das Empfehlungssystem macht von diesen Beweismitteln Gebrauch um sicherzustellen, dass jede Einheit Betrugsverhalten meldet, sobald sie solches wahrnimmt. **(2)** Soziale Beweismittel dienen dazu, beidseitige Vertrauensbeziehungen in Form von Bürgschaften sichtbar zu machen. Diese Bürgschaften werden in Selbstempfehlungen verbreitet. Der eigene Ansatz sorgt dafür, dass normative Einheiten durch die Angabe der eigenen Bürgen glaubwürdig ihren Typ signalisieren und sich somit von den strategischen Einheiten absetzen können.

Die simulative Evaluation des Ansatzes weist nach, dass die beiden Maximen des Entwurfs eingehalten werden und damit Betrugsverhalten effektiv eingedämmt wird. Das Schlüsselproblem der simulativen Evaluation ist es, alle Arten von Fehlverhalten, das die Einheiten unter realistischen Rahmenbedingungen zeigen, zu antizipieren. Zu diesem Zweck werden Versuchspersonen an einer interaktiven Simulation beteiligt. Ohne technische Vorkenntnisse zu benötigen, zeigen diese Versuchspersonen auf spielerische Weise viel versprechende Arten von Fehlverhalten auf. Mit Hilfe einiger Simulationswerkzeuge wird solches Fehlverhalten in der Gesamtevaluation des Ansatzes berücksichtigt. Für das Campus-Szenario zeigen die Simulationsergebnisse, dass die meisten Einheiten normativ am System teilnehmen. Mehrere Sensibilitätsanalysen bezüglich unterschiedlicher Parameter der Systemumgebung stellen sicher, dass dieses Resultat robust gegenüber abweichenden Rahmenbedingungen ist. Dadurch wird gezeigt, dass die These dieser Arbeit zutrifft. Die Vision von Informationssystemen wie dasjenige im Campus-Szenario lässt sich also realisieren.

1.5 Aufbau der Arbeit

Diese Dissertation ist folgendermaßen gegliedert:

- *Ausgangspunkt (Teil I):*

  In Kapitel 1 wurde das Campus-Szenario und die These dieser Arbeit vorgestellt. Der Stand der Forschung und Technik wird in Kapitel 2 behandelt. Es wird untersucht, wie existierende Ansätze Betrugsverhalten vermeiden und ob dazu Abstriche am Grad der Autonomie und der Selbstorganisation notwendig sind. Kapitel 3 gibt einen Überblick der Grundideen und Leistungen des eigenen Ansatzes.

- *Entwurf (Teil II):*

  Die Theorien anderer Forschungsbereiche, auf die der Entwurf des eigenen Ansatzes aufbaut, werden als Grundlagen in Kapitel 4 vorgestellt. Kapitel 5 erörtert die Beziehung zwischen dem Entwurf des Systems zur verteilten Vertrauensbildung und der Autonomie der am System teilnehmenden Einheiten. Daraus ergibt sich eine Typisierung der Einheiten und die Methodik des Systementwurfs. Kapitel 6 betrachtet lokale Vertrauensbildung, bei der lediglich eigene Transaktionserfahrungen einfließen. Damit wird das Fundament für die verteilte Vertrauensbildung gelegt, bei der Einheiten ihre Erfahrungen austauschen können. Zu diesem Zweck werden in Kapitel 7 transaktionale Beweismittel und in Kapitel 8 soziale Beweismittel vorgeschlagen. In beiden Kapiteln werden die Komponenten der Vertrauensbildung entsprechend erweitert.

- *Evaluation (Teil III):*

  In Kapitel 9 stelle ich eine Methodik zur simulativen Evaluation des eigenen Ansatzes vor. Dazu werden die Elemente und die Simulationswerkzeuge des Evaluationsprozesses erörtert. Dieser Prozess wird in Kapitel 10 durchlaufen. Als Ergebnis erhalten wir, unter welchen Rahmenbedingungen die These dieser Arbeit validiert werden kann.

- *Abschließende Betrachtungen (Teil IV):*

  Kapitel 11 fasst diese Arbeit zusammen. Nachfolgend werden in Kapitel 12 weiterführende Konzepte vorgestellt. Sie untersuchen, wie mit erschwerten Rahmenbedingungen umgegangen werden kann. Die Arbeit wird in Kapitel 13 mit einem Ausblick abgeschlossen.

Kapitel 2

知彼知己 百戦不殆

"Kenne die Lage deines Gegenübers und deiner selbst, und du wirst hundert Auseinandersetzungen unbeschadet überstehen."
(Sun-tze's Kunst der Kriegsführung, 3.18)

Stand der Forschung und Technik

Im letzten Kapitel ist die *Vision* von Informationssystemen, die wie im Campus-Szenario vollständig auf die Geräte der Endbenutzer verteilt sind, vorgestellt worden. Gemäß unserer These mutmaßen wir, dass sich diese Vision realisieren lässt. Dazu ist aber der Zielkonflikt zwischen Selbstorganisation des Informationssystems, Autonomie der teilnehmenden Einheiten und Effektivität der Betrugsvermeidung zu lösen.

Am Anfang dieses Kapitels wird ein kurzer Überblick über die Techniken der Automatisierung von Informationssystemen gegeben (Abschnitt 2.1). Sie ermöglichen eine Teilnahme der menschlichen Benutzer am Informationssystem im Sinne des Campus-Szenarios.

Im Rest dieses Kapitels untersuchen wir, ob bestehende Forschungsansätze für die Realisierung der Vision ausreichen. Dabei legen wir ein besonderes Augenmerk darauf, wie sehr die Dimensionen des Zielkonflikts berücksichtigt werden. Es zeigt sich, dass die Ansätze bei unterschiedlichen Dimensionen Kompromisse eingehen. Die Darstellung der Ansätze gliedert sich entsprechend dieser Ausrichtung:

- *Kompromisse bezüglich der Autonomie:* Teile des Informationssystems werden auf *manipulationssichere Hardware* platziert (Abschnitt 2.2). Betrug wird dadurch unmöglich gemacht.
- *Kompromisse bezüglich der Selbstorganisation:* Im Informationssystem sind *vertrauenswürdige Dritte* verfügbar (Abschnitt 2.3). Sie machen Betrugsverhalten unvorteilhaft, indem sie die Auswirkungen von Betrug rückgängig machen oder es erreichen, dass Betrüger von weiterer Kooperation ausgeschlossen werden.
- *Kein Kompromiss bezüglich Autonomie und Selbstorganisation:* Die Einheiten nehmen an einer *verteilten Vertrauensbildung* teil (Abschnitt 2.4). Sie erreicht es auf selbstorganisierende Art und Weise, dass Betrüger von weiterer Kooperation ausgeschlossen werden.

2.1 Grundlagen der Automatisierung

Im Campus-Szenario wurde ein Informationssystem beschrieben, das hochgradig *automatisiert* ist. Dies manifestiert sich darin, dass Kooperation nicht direkt zwischen menschlichen Benutzern

sondern zwischen ihren Einheiten stattfindet. Eine solche Automatisierung ist wegen der dadurch verbundenen Zeit- und Kostenersparnis wünschenswert.

Als Voraussetzung zur Automatisierung muss jeder menschlicher Benutzer seine Einheit entsprechend seiner Wünsche einstellen können. Darauf basierend muss seine Einheit in der Lage sein, im Sinne ihres menschlichen Prinzipals im Informationssystem zu handeln.

Die einzelnen Aspekte der Automatisierung werden von verschiedenen Forschungsbereichen behandelt. Ihre Techniken sind orthogonal zu dem Problemkreis dieser Arbeit, der effektiven Betrugsvermeidung, und können daher mit dem Ansatz dieser Arbeit kombiniert werden. Im Folgenden wird ein Überblick dieser Techniken der Automatisierung gegeben. Der Überblick ist kurz gehalten, da lediglich die prinzipielle Machbarkeit der angestrebten Automatisierung zu zeigen ist.

Einstellung von Benutzerpräferenzen. Eine Einheit kann nur dann im Sinne ihres menschlichen Prinzipals handeln, wenn sie seine *Präferenzen* kennt. Die Spezifikation der Präferenzen muss quantitativ sein, damit Entscheidungen zwischen verschiedenen Handlungsalternativen als einfache Nutzenmaximierung getroffen werden können [ECCC94]. Daher ist vom menschlichen Benutzer zu spezifizieren:

- *Nutzenstruktur:* Wie wertvoll sind für den Benutzer Informationen und die Erbringung von Informationsdiensten? Dementsprechend legt seine Einheit ihre Schwerpunkte im Informationssystem.

- *Kostenstruktur:* Die Handlungen einer Einheit ziehen Berechnungs-, Speicher- und Kommunikationsaufwand nach sich. Der menschliche Benutzer muss darlegen, welche Kosten dadurch für ihn entstehen.

Die quantitative Beschreibung der eigenen Präferenzen ist für einen menschlichen Benutzer eine komplexe und zeitraubende Aufgabe. Daher werden in der Literatur zu *Benutzeragenten* Ansätze vorgeschlagen, die aus den Benutzerentscheidungen Präferenzen ableiten können. Eine Übersicht dieser Ansätze wird in [JT05] gegeben. Die meisten von ihnen basieren darauf, ein *Benutzermodell* aufzustellen, das die kognitiven Prozesse der Benutzer erfasst. Die Parametrisierung des Modells wird durch Benutzerentscheidungen angepasst und spiegelt dadurch die Präferenzen des Benutzers wider.

Automatisierte Findung, Aushandlung und Nutzung von entfernten Informationen und Informationsdiensten. Weiß eine Einheit von den Wünschen ihres menschlichen Prinzipals, so setzt sie diese im Informationssystem um. Dazu bedarf es Techniken, wie benötigte Informationen und Informationsdienste aufgefunden werden und wie ihre Benutzung ausgehandelt und durchgeführt wird. Zur Beantwortung dieser Fragestellungen gehen wir im Folgenden auf die Techniken ein, die in der Literatur hierfür vorgeschlagen werden. Um unnötige Wiederholungen zu vermeiden, werden in der nachfolgenden Betrachtung Information und Informationsdienst unter dem Begriff Information zusammengefasst.

Um entfernte Informationen *auffinden* zu können, müssen sich die Einheiten des Informationssystems gegenseitig mitteilen, über welche Informationen sie jeweils verfügen. Eine einfache aber sehr aufwändige Lösung besteht darin, dass jede Einheit regelmäßig eine Liste der ihr verfügbaren Informationen an alle anderen erreichbaren Einheiten verteilt. Der Forschungsbereich der *Overlay-Netze* versucht diesen Aufwand zu vermindern, indem das Kommunikationsnetz der Einheiten

derart strukturiert wird, dass solche Listen nur an die strukturell benachbarten Einheiten verteilt werden müssen. Verschiedenartig strukturierte Overlay-Netze werden in [RFH+01, KKRO03] vorgestellt.

Nach dem Auffinden einer Einheit, die über die gewünschte Information verfügt, ist mit ihr *auszuhandeln*, welche Gegenleistung für die Bereitstellung dieser Information erbracht werden muss. Als Gegenleistung bietet sich eine Information an, die von dieser Einheit benötigt wird. Die *Theorie des Feilschens* (engl.: bargaining theory) befasst sich mit der Frage, wie ein solches Aushandeln erfolgen muss. Eine Übersicht verschiedener Ansätze zum *bilateralen* Feilschen wird in [Usu02] gegeben. Sie geben jeweils vor, wie eine Einheit gegenüber einen potentiellen Transaktionspartner entsprechend ihrer jeweiligen Präferenzen verhandeln muss, damit in der nachfolgenden Transaktion ein größtmöglicher Nutzen für sie entsteht. Bei Informationsdiensten kann es aufgrund von Ressourcenbeschränkungen seitens des Anbieters dazu kommen, dass verschiedene Einheiten um die Benutzung eines Informationsdienstes konkurrieren. In einem solchen Fall ist der Einsatz von *Auktionen* sinnvoll, wie eine Übersicht verschiedener Auktionsarten zeigt [BSGA01]. Der Zusammenhang zwischen solchen Auktionen und der Auffindung von Informationen in Overlay-Netzen wird in [OK03a] aufgezeigt.

Ist das Aushandeln erfolgreich, so kommt es zum Austausch der Informationen zwischen den beiden Transaktionspartnern. Bei der automatisierten *Benutzung* von Informationsdiensten ergibt sich hierbei das Problem, dass beide Einheiten dieselben Vorstellungen bezüglich der Schnittstelle und Funktionsweise des Dienstes haben müssen. Zur Lösung dieses Problems ist ein Ansatz anzuwenden, der die automatisierte Ausführung von Informationsdiensten unterstützt. Einen solchen Ansatz stellen die *Web-Services* [W3C04] dar, die in letzter Zeit weite Verbreitung gefunden haben.

Automatisierte Einschätzung menschlicher Fähigkeiten. Manche der Informationen, die im Informationssystem verfügbar gemacht werden, sind von den menschlichen Benutzern selbst erstellt. Im Campus-Szenario fasst zum Beispiel Anna ihr Verständnis der Vorlesung in einem Vorlesungsmitschrieb zusammen. Es liegt auf der Hand, dass die Qualität der selbst erstellten Informationen von den *Fähigkeiten* der menschlichen Benutzer abhängt. Für das Fällen von Transaktionsentscheidungen muss eine Einheit daher die erwartete Qualität der angeforderten Informationen *einschätzen* können. Durch die Zielvorgabe der Automatisierung sollen solche Einschätzungen möglichst ohne Rücksprache mit dem menschlichen Prinzipal gemacht werden.

Grundlage für die Einschätzung der zu erwartenden Qualität von Informationen muss die Qualität bisher erhaltener Informationen sein. So sind zum Beispiel im Campus-Szenario Bob und Claude besonders an Annas Vorlesungsmitschrieb interessiert, da sie mit ihren bisherigen Mitschrieben zufrieden waren. Die *Bewertung* der Qualität erhaltener Informationen übersteigt jedoch die technischen Möglichkeiten der Einheiten und muss daher manuell durchgeführt werden. Im Campus-Szenario bedeutet dies zum Beispiel, dass Bob Annas Vorlesungsmitschrieb zum Lernen benutzt und anschließend seinem Gerät eine Bewertung seiner Qualität abgibt[1].

Der Forschungsbereich der *Empfehler-Systeme* (engl.: recommender systems) befasst sich damit, wie aus den Bewertungen auf die Fähigkeiten menschlicher Benutzer gefolgert werden kann. Dazu wird angenommen, dass menschliche Benutzer bei einer ähnlichen Art von Information auch ähnliche Fähigkeiten besitzen. Im Campus-Szenario bedeutet dies, dass aus Annas guten Vorlesungsmitschrieben gefolgert werden kann, dass sie zum Beispiel auch gute Prüfungsprotokolle schreiben kann. Ein Empfehler-System, das diesen Ansatz verfolgt, wird in [Vid03, Vid05]

[1] Da Bob dadurch seinem Gerät auch seine Präferenz bezüglich der Autoren von Vorlesungsmitschrieben mitteilt, entspricht dies im weiteren Sinne der Einstellung der Benutzerpräferenzen.

vorgeschlagen. Es beinhaltet die Möglichkeit, auch die Bewertungen anderer Benutzer einzusehen. Damit wird die Informationsbasis für die Einschätzung menschlicher Fähigkeiten erweitert, wodurch die Genauigkeit der Einschätzung weiter gesteigert wird.

Automatisierte Authentifizierung und Nichtabstreitbarkeit. Jeder menschliche Benutzer wird von seiner Einheit im Informationssystem vertreten. Es bedarf daher einer Technik, mit der sich eine Einheit auf glaubwürdige Weise als Repräsentant ihres menschlichen Prinzipals ausweisen kann. Mit Hilfe einer solchen Technik können zwei Annahmen umgesetzt werden, die im Campus-Szenario gemacht werden:

- *Authentifizierung:* Miteinander kommunizierende Einheiten erkennen gegenseitig ihre Identität. Es ist also zum Beispiel für Manuels Einheit nicht möglich, sich als Annas Einheit auszugeben.

- *Autorenschaft:* Erstellt ein menschlicher Benutzer eine Information, so versieht seine Einheit die Information mit einem Vermerk, dass sie von diesem Benutzer stammt. Damit ist ausgeschlossen, dass ein Benutzer wie Manuel einen Vorlesungsmitschrieb im Namen von Anna erstellt und dadurch bessere Gelegenheiten zu Transaktionen erhält.

Aus der Kryptographie ist die *Infrastruktur mit öffentlichen Schlüsseln* (engl.: public key infrastructure; *PKI*) als Technik bekannt, die diese Forderungen erfüllt [Kou03]. Ihr Grundprinzip ist wie folgt: Jede Einheit besitzt einen öffentlichen und einen privaten Schlüssel. Den privaten Schlüssel behält sie für sich, während sie den öffentlichen Schlüssel ihren Kommunikationspartnern weitergibt. Da die Schlüssel mit Hilfe von Algorithmen erstellt sind, deren Umkehrung sehr aufwändig ist[2], kann aus dem öffentlichen Schlüssel nicht der private Schlüssel abgeleitet werden. Signiert eine Einheit mit ihrem privaten Schlüssel Nachrichten oder Informationen im Binärcode, kann jede andere Einheit mit Hilfe des dazugehörigen öffentlichen Schlüssels überprüfen, dass die Signatur von dieser Einheit stammt.

Um die PKI als Mittel zur Authentifizierung und Autorenschaft zu gebrauchen, sind die folgenden Schritte notwendig: Vor dem Einsatz des Informationssystems kreiert der Betreiber des Informationssystems ein Schlüsselpaar aus öffentlichem und privatem Schlüssel. Wenn ein menschlicher Benutzer mit seinem Gerät am Informationssystem teilnehmen möchte, so erhält er initial außer der Systemsoftware auch Folgendes vom Betreiber[3]:

- Eine Identität, zum Beispiel als Pseudonym. Der Betreiber muss darauf achten, dass jeder Benutzer nur eine Identität erhalten kann. Im Campus-Szenario ist diese Anforderung erfüllt, da die Universität die Identität der Studenten für das Ausstellen von Studentenausweisen bereits überprüft hat.

- Ein Schlüsselpaar aus öffentlichem und privaten Schlüssel.

- Ein Zertifikat des Betreibers, dass die Identität und der öffentliche Schlüssel des Benutzers zusammengehören. Technisch gesehen signiert dazu der Betreiber mit seinem privaten Schlüssel die Identität und den öffentlichen Schlüssel des Benutzers.

[2]Zum Einsatz kommt in der PKI vor allem das Problem, das Produkt zweier großer Primzahlen zu faktorisieren.

[3]Die Vergabe der Systemsoftware, der Identität des Benutzers und der kryptographischen Marken kann auf unterschiedliche Arten geschehen. Im Campus-Szenario bietet sich dafür eine Einrichtung wie das Rechenzentrum, das Studentenwerk oder die Universitätsbibliothek an. Auch die Vergabe über das Internet ist denkbar, wenn die Benutzer ein Mittel zur Authentifizierung dort haben. Hierfür kann eine Standardsoftware wie die Software-Tankstelle [LO03] eingesetzt werden.

- Der öffentliche Schlüssel des Betreibers. Dieser wird benötigt um die Zertifikate anderer Benutzer überprüfen zu können.

Unter diesen Voraussetzungen ist jede Einheit in der Lage, beliebige Informationen zu signieren. Eine solche Signatur ist insofern *nicht-abstreitbar* [ISO97], als andere Einheiten nach Erhalt der signierten Information wissen, wer sie ausgestellt hat. In diesem Sinne stellt die Technik der PKI ein Mittel zur Authentifizierung und Autorenschaft zur Verfügung.

Abschließend ist noch zu klären, wie eine Einheit Zugang zu einem benötigten Zertifikat erhält. Die klassischen Ansätze der PKI benutzen hierfür zentrale Server [Kou03], die aufgrund der Selbstorganisation der hier betrachteten Informationssysteme jedoch ausscheiden. Eine triviale Lösung des Problems besteht darin, dass an jede signierte Information das Zertifikat ihres Ausstellers angehängt wird [Obr04a]. Damit wird sichergestellt, dass jede Einheit, die die Identität des Ausstellers einer Information überprüfen muss, Zugang zu seinem Zertifikat besitzt. Dieser Ansatz kann effizienter gemacht werden, indem Zertifikate nur auf Anfrage weiter gegeben werden. Dass dies auch in stark partitionierenden Netzen wie im Campus-Szenario möglich ist, zeigt folgende Betrachtung: Sei eine signierte Information von Einheit A ausgestellt worden. Sie wird nun von Einheit B an Einheit C weiter gegeben. Wenn Einheit C das Zertifikat von Einheit A nicht kennt, kann sie es bei Einheit B anfordern. Einheit B besitzt dieses Zertifikat, da ursprünglich auch sie die Signatur der Information überprüft hat. Damit lässt sich ein induktives Argument anwenden mit der Induktionsbasis, dass jede Einheit ihr eigenes Zertifikat kennt. Wir erhalten daher, dass ein benötigtes Zertifikat verfügbar ist, auch wenn die jeweilige Einheit nicht erreichbar sein sollte. Der Zugang zu Zertifikaten stellt also trotz der Selbstorganisation des Informationssystems kein Problem dar.

2.2 Systeme mit manipulationssicherer Hardware

Ein wesentlicher Teil unserer Vision von verteilten Informationssystemen liegt darin, dass jeder menschliche Benutzer mit seinem Informationsgerät am System teilnehmen kann. Der Benutzer und sein Gerät sind autonom gegenüber dem Informationssystem, da der Benutzer sein Gerät kontrolliert. Dadurch entsteht die Möglichkeit der *Manipulation* der Systemsoftware und letztendlich Betrugsverhalten im Informationssystem. Wenn also dem menschlichen Benutzer die Kontrolle über sein Gerät entzogen wird, lässt sich Betrugsverhalten von Vornherein verhindern.

Ein Ansatz besteht darin, die Systemsoftware auf *manipulationssicherer Hardware* des Gerätes des menschlichen Benutzers zu platzieren. In diesem Abschnitt untersuchen wir, ob dieser Ansatz tauglich ist und, wenn ja, welche Kosten für die Teilnahme am Informationssystem durch die Inbetriebnahme manipulationssicherer Hardware entstehen. Wir unterscheiden zwei Möglichkeiten zur Erzielung von Manipulationssicherheit:

- Die Systemsoftware des Informationssystems wird fest auf ein *speziell* für das Informationssystem gefertigtes Hardware-Modul abgelegt. Zur Teilnahme muss ein menschlicher Benutzer ein solches Hardware-Modul erwerben.

- Der Benutzer ist im Besitz von manipulationssicherer *Mehrzweck*-Hardware. Damit reicht zur Teilnahme die Installation der Systemsoftware auf dieser Hardware aus.

Beide Möglichkeiten zielen auf die Gewährleistung davon, dass die Teilnahme am Informationssystem nur mit der originalen Systemsoftware erfolgen kann. Dadurch dürfen in der Systemsoftware Verhaltensweisen festgelegt werden, deren Befolgung für die jeweilige Einheit zwar von

Nachteil, für das Gesamtsystem jedoch von Vorteil ist. Solches Verhalten nennen wir im Folgenden *individuell irrational*. Ein Beispiel hierfür ist der Verzicht auf Betrugsverhalten. Individuell rationales Verhalten wird also nur deswegen nicht gezeigt, weil die Benutzer keine Möglichkeit zu der dafür notwendigen Manipulation der Systemsoftware erhalten.

Im Folgenden gehen wir auf die zwei Möglichkeiten zur Erzielung von Manipulationssicherheit ein. Abschnitt 2.2.1 bespricht spezialisierte Hardware, während in Abschnitt 2.2.2 auf Mehrzweck-Hardware eingegangen wird. Abschließend wird die Tauglichkeit der Ansätze in Abschnitt 2.2.3 bewertet.

2.2.1 Spezialisierte Hardware

Ein trivialer Ansatz besteht darin, dass alle Teile des Informationsgeräts manipulationssicher sind. Im Campus-Szenario würde das zum Beispiel bedeuten, dass jeder Student, der am System teilnehmen will, ein eigens dafür hergestelltes Informationsgerät kaufen müsste. Dies steht im direkten Widerspruch zu der Vision, dass menschliche Benutzer mit ihren unabhängig erworbenen Informationsgeräten am System teilnehmen können. Die anfallenden Kosten stellen eine zu hohe Eintrittsbarriere dar.

Daher wird in TermiNodes Projekt [BH03] vorgeschlagen, dass Informationsgeräte der menschlichen Benutzer mit einer Zusatzhardware, die manipulationssicher ist, ausgestattet werden. Der Erwerb und die Installation dieses Hardware-Moduls ermöglicht dann die Teilnahme am System. Um die Anforderungen an die Ressourcen des Hardware-Moduls und damit dessen Kosten zu minimieren, werden nur besonders sensible Teile der Systemsoftware auf dieses Hardware-Modul platziert. Insbesondere bleiben die Netzwerkschnittstelle und große Teile des benutzten Speichers manipulierbar. Durch den Einsatz geeigneter kryptographischer Verfahren wird jedoch gezeigt, dass eine solche Manipulation für den menschlichen Benutzer nicht vorteilhaft ist [But01]. Im Kern liegt das daran, dass die Teile der Systemsoftware, deren Manipulation zu Betrugsverhalten führen kann, auf dem manipulationssicheren Hardware-Modul platziert sind. Im Vergleich zum trivialen Ansatz der vollständig manipulationssicheren Informationsgeräte führt dieser Ansatz zwar zu geringeren Ausstattungskosten. Dies ändert aber nichts daran, dass der menschliche Benutzer Zusatzhardware erwerben muss, bevor er am System teilnehmen kann. Auch aber nicht nur im Campus-Szenario zeigt sich, dass dies zumindest ein großes Hindernis für die Verbreitung und den Erfolg des Informationssystems ist.

2.2.2 Mehrzweck-Hardware

Die Lage ändert sich, wenn die Informationsgeräte der menschlichen Benutzer von sich aus Manipulationssicherheit gewährleisten. In diesem Falle stellt das Informationsgerät eine manipulationssichere Mehrzweck-Hardware dar. Für die Teilnahme am System muss also lediglich die Systemsoftware auf diese Mehrzweck-Hardware installiert werden. Um die Installation einer manipulierten Version der Systemsoftware zu verhindern, müssen zudem zwei Forderungen erfüllt sein: **(1)** Die Mehrzweck-Hardware erlaubt nur die Verwendung von Software, die als nicht manipuliert zertifiziert ist. **(2)** Nur die Originalversion der Systemsoftware erhält ein solches Zertifikat.

Kann der Systementwerfer erwarten, dass die Informationsgeräte der menschlichen Benutzer Manipulationssicherheit gewährleisten? Die zurzeit erhältlichen und im Einsatz befindlichen Informationsgeräte können dies nicht. Die *Trusted Computing Group*[4] (TCG) [TCP00] hat zum Ziel, ebendies zu ändern. Im Folgenden gehen wir daher auf ihren Ansatz ein.

[4]Die TCG nannte sich früher *Trusted Computing Platform Alliance*.

Die Grundidee der TCG entstammt aus der Veröffentlichung von Arbaugh et al. [AFS97]: Als Voraussetzung wird die zu verwendende Hardware und Software (inklusive Betriebssystem) von einer *Zertifizierungsauthorität* zertifiziert. Im Laufe des Boot-Prozesses achtet ein Chip (so genannter *Fritz-Chip*) des Informationsgeräts darauf, dass nur zertifizierte Hardware und Software verwendet werden. Auch nach Abschluss des Boot-Prozesses ist das Aufspielen von manipulierter Software unmöglich, da ihr das notwendige Zertifikat fehlt und das Betriebssystem damit ihre Verwendung verweigert. In einer abgewandelten Version dieser Technik nimmt der Fritz-Chip nur eine passive Rolle ein [And03b]. Es ist zwar möglich, dass nicht-zertifizierte und damit potentiell manipulierte Software auf dem Gerät verwendet wird. Jedoch erhält nur zertifizierte Software vom Fritz-Chip Unterstützung dafür, sich anderen Geräten gegenüber als zertifiziert auszuweisen. Wenn unsere Systemsoftware also entsprechend zertifiziert ist, können sich die Informationsgeräte mit der Originalversion der Systemsoftware eindeutig untereinander erkennen. Manipulierte Versionen können damit nicht am Informationssystem teilnehmen.

Der Ansatz der TCG schränkt jedoch unsere Vision von Informationssystemen, an denen jeder menschliche Benutzer mit seinem Informationsgerät teilnehmen kann, ein. Das liegt daran, dass zur Teilnahme am Informationssystem das Gerät den Standard der TCG umsetzen müsste. Auf absehbare Zeit hinaus ist jedoch unter anderem durch den Einsatz älterer Geräte zu erwarten, dass nicht jedes Informationsgerät dieser Anforderung genügt. Das hat zur Folge, dass ein menschlicher Benutzer ein entsprechendes Gerät erwerben müsste, um am System teilzunehmen. Diese Eintrittsbarriere hält nicht nur im Campus-Szenario menschliche Benutzer von der Teilnahme am System ab.

Abgesehen von dieser Einschränkung gibt es eine Reihe von Gründen gegen den Ansatz der TCG. Im Folgenden werden diese Gründe vorgebracht. Sie lassen es als zumindest fragwürdig erscheinen, ob sich der Ansatz der TCG in der Praxis durchsetzt.

Die im Ansatz vorgesehene technische Umsetzung kann *nicht effektiv* Manipulationssicherheit gewährleisten. Eine Reihe von Angriffspunkten wird in [Hei02] aufgezeigt. Sie rühren einerseits von inhärenten Problemen der Umsetzung von Manipulationssicherheit [AK96]. Andererseits sind auch Schwächen im Entwurf der TCG zu erkennen. So bietet zum Beispiel die ungesicherte Kommunikation zwischen CPU und Fritz-Chip Möglichkeiten zur Manipulation.

Das Ziel der TCG ist es, durch ihren Ansatz einen Mehrwert für den Endbenutzer zu schaffen. Die Analyse von Anderson [And03b] zeigt jedoch, dass *kein* solcher *Mehrwert* für den Endbenutzer zu erwarten ist. Ein besonders heikler Punkt wird in [And03c] herausgestellt: Durch die Zertifizierung von Software verspricht die TCG dem Endbenutzer, dass sein Gerät sicher vor Viren ist. Jedoch ist das vorgesehene Zertifizierungsverfahren *EAL3* nicht streng genug, um dies zu gewährleisten. Es ist also zu erwarten, dass selbst zertifizierte Software Angriffspunkte bietet und ihr Verhalten somit doch manipuliert werden kann. Durch das Fehlen eines Mehrwertes für den Endbenutzer ist für die TCG keine Unterstützung seitens der Benutzer-Verbände zu erwarten.

Ein weiteres Hindernis ist *rechtlicher* Art. Die TCG als Standardisierungsgruppe verstößt nämlich nach Ansicht von Rechtsexperten gegen das europäische Wettbewerbsrecht [KN03]. Genauer gesagt werden von der TCG drei Forderungen der Paragraphen § 81 und § 82 des Kartellgesetzes nicht erfüllt. Sie verlangen, dass Standards *nicht-bindend*, *offen* und *nicht-diskriminierend* sind. Diese Forderungen werden aus drei Gründen verletzt: **(1)** Die Bedingungen der Mitgliedschaft an der TCG benachteiligen kleine und mittelständische Unternehmen. Dies liegt an den hohen Beitrittsgebühren für die Teilnahme an der TCG. **(2)** Die Mitgliedschaft ist von zentraler Bedeutung, um Einfluss auf die Standardisierung zu nehmen, technisches Wissen zu erlangen, frühen Zugriff auf Spezifikationen zu haben und wesentliche Patente der Plattform lizensiert zu bekommen. **(3)** Die Zertifizierung von Software nach EAL3 ist so kostspielig, dass Software klei-

ner Unternehmen und Open-Source-Software nicht zertifiziert werden kann [And03c]. Es handelt sich also um einen diskriminierenden Standard.

Abgesehen der rechtlichen Probleme bestehen weitere Vorbehalte *politischer* und *ideologischer* Art gegen die TCG [And03c]. Sie rühren daher, dass den Endbenutzern die Kontrolle über ihre Geräte entzogen wird. Sie wird nunmehr von der Zertifizierungsautorität ausgeübt. Auch wenn die Einzelheiten nicht im Standard der TCG festgelegt sind, lässt sich durch die massive Präsenz US-amerikanischer Unternehmen in der TCG absehen, dass die Zertifizierungsautorität unter der indirekten Kontrolle nur weniger Länder stehen wird. Für die Regierungen der nicht beteiligten Länder ist es daher unter Umständen im nationalen Interesse, gegen die Bemühungen der TCG vorzugehen. Auch aus der Sicht der Endbenutzer gibt es Vorbehalte gegen die TCG. Die Aussicht darauf, dass die eigenen Geräte im Sinne der TCG entfernt kontrolliert werden könnten, hat bereits zu Protesten geführt[5].

Abschließend halten wir fest, dass der Ansatz manipulationssicherer Mehrzweck-Hardware wie durch die TCG verfolgt nicht gangbar ist.

2.2.3 Bewertung

Systeme mit manipulationssicherer Hardware entziehen dem menschlichen Benutzer Kontrolle über sein Informationsgerät. Durch diese Einschränkung der Autonomie der teilnehmenden Einheiten lässt sich verhindern, dass manipulierte Versionen der Systemsoftware benutzt werden. Damit lässt sich in der originalen Systemsoftware individuell irrationales Verhalten wie der Verzicht auf Betrugsverhalten durchsetzen.

Dieser Ansatz steht und fällt mit der Verfügbarkeit manipulationssicherer Hardware für den Endbenutzer. Die Untersuchung zeigt, dass die Möglichkeiten hierzu entweder (bei spezialisierter Hardware) dem Endbenutzer Kosten verursachen, die als zu hohe Eintrittsbarriere wirken, oder (bei Mehrzweck-Hardware) schlichtweg nicht gangbar sind. Als Ergebnis erhalten wir, dass es unbedingt erforderlich ist, den Teilnehmern am Informationssystem Autonomie zu gewähren. Damit können nur solche Ansätze zum Erfolg führen, die damit zurechtkommen, dass auf den Geräten mancher Teilnehmer manipulierte Versionen der Systemsoftware installiert sind.

2.3 Systeme mit vertrauenswürdigen Dritten

Im Campus-Szenario ist ein Informationssystem beschrieben, das vollständig auf die Geräte der menschlichen Teilnehmer verteilt ist. Dadurch, dass jeder Benutzer und sein Gerät autonom gegenüber dem Informationssystem sind, entsteht die Möglichkeit zur Manipulation und folglich zu Betrugsverhalten. Um das zu verhindern, muss das Informationssystem Einfluss auf das Verhalten der teilnehmenden Einheiten ausüben können.

Ein Ansatz besteht also darin, dass das Informationssystem um eine zentrale Instanz erweitert wird, die im System regulierend eingreifen kann. Da die zentrale Instanz vom Betreiber des Informationssystems kontrolliert wird, handelt es sich bei ihr um einen *vertrauenswürdigen Dritten*. Ein regulierender Eingriff ist nötig, wenn es durch Betrugsverhalten zu einem Konflikt zwischen zwei Transaktionspartnern kommt. In der Literatur sind dazu folgende Klassen von Ansätzen vorgeschlagen worden:

- *Konfliktvermeidung:* Der vertrauenswürdige Dritte schaltet sich derart in den Ablauf einer Transaktion ein, dass es für Einheiten unmöglich ist, zu betrügen. Dadurch werden Konflikte

[5] Der Protest hat sich zum Beispiel in Internetseiten wie *http://www.againsttcpa.com* manifestiert.

zwischen Einheiten *vermieden*. Ansätze hierzu werden in Abschnitt 2.3.1 vorgestellt.

- *Konfliktlösung:* Nach dem Auftreten eines Konflikts schaltet sich der vertrauenswürdige Dritte ein, um die Vorteile des Betrugsverhaltens rückgängig zu machen. In dieser Hinsicht werden Konflikte im Nachhinein *gelöst*, so dass Betrugsverhalten unvorteilhaft wird. Abschnitt 2.3.2 geht auf solche Ansätze ein.

- *Bewertung und Verbreitung von Reputation:* Der vertrauenswürdige Dritte ist nicht in der Lage, Konflikte im Vorfeld zu vermeiden oder nach ihrem Auftreten zu lösen. Hingegen sammelt der Dritte Informationen über das vergangene Verhalten der Einheiten, um einen Anhaltspunkt über ihr zukünftiges Verhalten zu erhalten. Der Dritte *bewertet* dazu die *Reputation* einer jeden Einheit und *verteilt* sie an interessierte Einheiten. Dadurch wird erreicht, dass betrügende Einheiten eine schlechte Reputation tragen und damit seltener als Transaktionspartner angenommen werden. Solchen Einheiten entgeht der Mehrwert, der aus der Teilnahme an Transaktionen entsteht, so dass ein negativer Anreiz für betrügerisches Verhalten gegeben wird. Ansätze hierzu werden in Abschnitt 2.3.3 vorgestellt.

Die besprochenen Ansätze werden abschließend in Abschnitt 2.3.4 bewertet.

2.3.1 Dritte zur Konfliktvermeidung

Konflikte zwischen Transaktionspartnern lassen sich von Vornherein vermeiden, wenn dafür gesorgt wird, dass Betrugsverhalten *nicht effektiv* sein kann. In der Literatur sind einige Ansätze bekannt, die dieses Ziel verfolgen. Sie schlagen dazu Protokolle vor, mit Hilfe derer Informationen zwischen den Transaktionspartnern ausgetauscht werden. Solche *Austauschprotokolle* stellen insofern *Fairness* zwischen den Transaktionspartner her, als entweder beide oder keiner der Transaktionspartner an die gewünschten Informationen gelangt [Aso98].

Im Folgenden wird auf diese Ansätze und ihre Austauschprotokolle genauer eingegangen. Dabei wird unterschieden zwischen **(1)** *pessimistischen* Protokollen, in denen der vertrauenswürdige Dritte zwischen den Transaktionspartner zwischengeschaltet ist, und **(2)** *Konten-basierten* Protokollen, in denen der Dritte im Hintergrund durch die Verwaltung der Teilnehmerkonten Fairness herstellt. Anhand dieser beiden Protokollarten werden die existierenden Ansätze am Ende dieses Abschnitts eingeordnet.

Entsprechend unseres Systemmodells aus Abschnitt 1.2.2 fasst die Besprechung die Bereitstellung von Informationen und das Erbringen von Informationsdiensten unter dem Begriff *Ausführen einer Aktion* zusammen. Unter dem Ergebnis einer Aktion verstehen wir also die bereitgestellte Information oder das Resultat des Informationsdiensts.

Pessimistische Austauschprotokolle. In einem pessimistischen Austauschprotokoll kommunizieren die Transaktionspartner über einen Dritten. Die Schritte eines solchen Protokolls sind in Abbildung 2.1 dargestellt. Sie lassen sich in drei Phasen unterteilen:

1. Die Einheiten A und B führen ihre Aktion aus und senden ihr Ergebnis an den Dritten. Zudem wird eine Spezifikation darüber übermittelt, welche Aktion vom jeweiligen Transaktionspartner erwartet wird.

2. Der Dritte überprüft, ob die beiden ausgeführten Aktionen den jeweiligen Spezifikationen entsprechen. Wurde eine Aktion nicht entsprechend der Spezifikation oder gar nicht ausgeführt, so bricht das Protokoll ab.

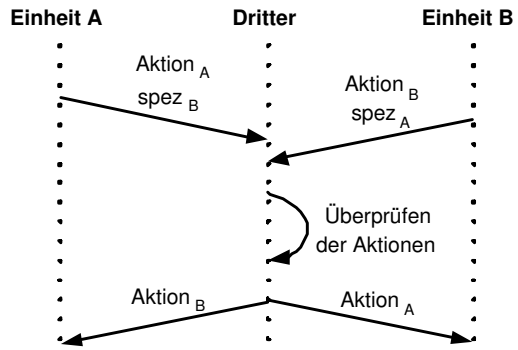

Abbildung 2.1: Ablauf eines pessimistischen Austauschprotokolls

3. Der Dritte leitet das Ergebnis der Aktionen an die Einheiten zurück.

Das Protokoll stellt sicher, dass eine Einheit aus Betrugsverhalten keinen Vorteil ziehen kann. Unterlässt nämlich eine Einheit das Ausführen ihrer Aktion, so bricht das Protokoll in der zweiten Phase ab und die betrügende Einheit erhält keinen Vorteil aus der Aktion des Transaktionspartners.

Konten-basierte Austauschprotokolle. Eine Alternative zu den pessimistischen Austauschprotokollen stellen Konten-basierte Protokolle dar. Jede Einheit erhält ein *Konto*, das von dem vertrauenswürdigen Dritten verwaltet wird. Kommt es zu einer Transaktion, so wird als Gegenleistung keine Aktion ausgeführt sondern ein *Scheck* ausgestellt und an den Transaktionspartner übermittelt [ON03]. Nach der Einlösung des Schecks beim Dritten werden die Konten der Transaktionspartner entsprechend angepasst.

Um Konflikte vermeiden zu können, muss der Ablauf von Konten-basierten Protokollen entsprechend Abbildung 2.2 wie folgt sein:

1. Vor der Transaktion mit Einheit B lässt sich Einheit A einen *Scheck* durch den Dritten ausstellen. Die Höhe des Schecks haben die beiden Einheiten vor der Transaktion untereinander ausgehandelt. Er richtet sich nach dem Aufwand von B für das Ausführen der Aktion und dem Nutzen aus der Aktion für A.

2. Einheit A überreicht den Scheck an Einheit B.

3. Einheit B führt die Aktion für Einheit A aus.

4. Einheit A gibt eine nicht-abstreitbare *Bestätigung*, dass B die Aktion ausgeführt hat und damit das Recht hat, den Scheck bei dem Dritten einzulösen.

5. Einheit B löst den Scheck beim Dritten ein und muss dafür die Bestätigung vorzeigen. Die Höhe des Schecks wird von A's Konto abgezogen und B's Konto zugeschrieben.

   - Besitzt B keine Bestätigung von A, so wird beim Einlösen des Schecks lediglich A's Kontostand vermindert.

Offensichtlich spielt A's Bestätigung eine maßgebliche Rolle bei diesem Protokoll:

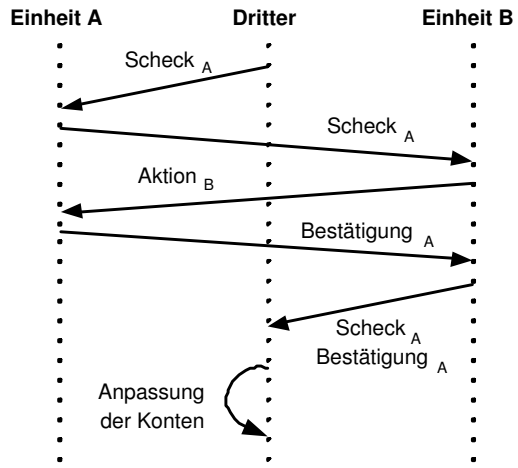

Abbildung 2.2: Ablauf eines Konten-basierten Austauschprotokolls

- Betrügt B, indem er den dritten Schritt nicht ausführt, so bringt ihm dies keinen Vorteil. Da ihm dann A's Bestätigung fehlt, wird der Scheck nämlich nicht auf seinem Konto gutgeschrieben.
- Betrügt A, indem er keine Bestätigung ausstellt, so ergibt sich für ihn keinen Vorteil, da auch ohne die Bestätigung sein Kontostand durch B's Einlösen des Schecks vermindert wird. Allerdings ergibt eine genaue Analyse, dass dieses Argument sehr fragwürdig ist:
  - Das Ausstellen und Übertragen der Bestätigung verursacht für A einen Aufwand. Einheit A ist daher nicht indifferent gegenüber der Bestätigung. Jedoch basiert die Effektivität dieses Protokolls auf eben diese Indifferenz.
  - Erhält Einheit B keine Bestätigung, so hat sie keinen Anreiz, dennoch den Scheck einzulösen, da er ihrem Konto nicht gutgeschrieben wird. Daher ist zu erwarten, dass die Bestrafung für eine betrügende Einheit A unterbleibt. Damit entsteht ein substanzieller Anreiz für A, die Bestätigung nicht auszustellen.

Der *Vorteil* eines Konten-basierten Protokolls gegenüber dem pessimistischen Protokoll ergibt sich daraus, dass das Ausstellen und Einlösen der Schecks zur Transaktion zeitlich versetzt geschehen kann. Dadurch muss der Dritte nicht zum Zeitpunkt der Transaktion von den Transaktionspartnern erreichbar sein. Ein weiterer Vorteil ergibt sich daraus, dass der Dritte nicht die Richtigkeit der Aktion anhand einer Spezifikation überprüfen muss. Allerdings werden diese Vorteile durch den *Nachteil* erkauft, dass das Protokoll nicht robust gegen Betrug ist. Daher sind pessimistische Austauschprotokolle den Konten-basierten vorzuziehen.

Einordnung der existierenden Ansätze. Das *pessimistische* Austauschprotokoll wird in [PSW98] besprochen und mit den Protokollen zur Konfliktlösung, die im nächsten Abschnitt besprochen werden, verglichen.

Das *Konten-basierte* Austauschprotokoll, wie es in diesem Abschnitt dargestellt wurde, entstammt aus [But01]. Es wird im *TermiNodes* Projekt eingesetzt, um Betrugsverhalten beim Weiterleiten von Nachrichten in einem Ad-hoc Netz zu unterbinden [BJHS03, JHB03, SBHJ03]. Eine Besonderheit dieses Ansatzes ist es, dass der vertrauenswürdige Dritte nicht auf einen entfernten

Server platziert wird. Stattdessen ist er Teil der Systemsoftware, die auf die Informationsgeräte verteilt wird. Um dennoch die Vertrauenswürdigkeit des Dritten garantieren zu können, geht TermiNodes von Manipulationssicherheit aus. Daher stellt der Ansatz einen Mittelweg zwischen Ansätzen mit manipulationssicherer Hardware und mit Dritten dar.

Eine *Variation* des Konten-basierten Protokolls wird in [Aso98] eingeführt. Sie besteht darin, dass Schecks von den Einheiten selbst und nicht vom Dritten ausgestellt werden. Allerdings wird der Empfänger eines Schecks dadurch genötigt, sich beim Dritten von der Deckung des Schecks zu vergewissern. Ist eine solche Nachfrage beim Dritten nur zeitlich versetzt möglich, so muss der Aussteller nicht gedeckter Schecks vom Dritten bestraft werden können. Dadurch werden Konflikte zwar nicht vermieden, sie werden aber im Nachhinein gelöst.

Eine Reihe von Ansätzen nimmt ein Konten-basiertes Protokoll an, um die Bereitstellung von Informationen und die Erbringung von Diensten zu belohnen. Diese Ansätze finden sich in [GLBML01, CA02, SAJ02]. Ein besonderes Augenmerk wird darauf gelegt, welche Höhe der Schecks als angemessen gelten darf. Hierfür gehen die Ansätze von verschiedenen Ausgangspunkten aus. Der Ansatz aus [BSGA01] basiert auf *selbstorganisierendem Feilschen*, wie es in Abschnitt 2.1 besprochen wurde. Einen weiteren Ansatzpunkt stellt der *Mechanismus-Entwurf* (engl.: mechanism design) dar, in dem der Dritte die Präferenzen der Transaktionspartner erfragt und darauf basierend selbst die Höhe des Schecks festlegt [KDMT00]. Ein verteilter Ansatz zum Mechanismus-Entwurf wird in [BS99, Bra01] vorgestellt. Ein weiterer Ansatz geht von der Theorie der öffentlichen Güter aus, indem er die im Informationssystem bereitgestellte Informationen als ein Allgemeingut ansieht [ST97]. All diese Ansätze haben gemein, dass sie eine angemessene Höhe der Schecks zu finden versuchen. Die Probleme des zugrunde liegenden Konten-basierten Protokolls werden dadurch jedoch nicht gelöst.

Eine grundlegende Variation zum Konten-basierten Austauschprotokoll wird in [Syv98] vorgeschlagen. An den Schritten 2 bis 4 des Protokolls wird festgehalten, jedoch wird die Verwaltung der Konten und damit der Dritte selbst überflüssig gemacht. Allerdings ist der letzte Schritt von Einheit A weiterhin sehr fragwürdig, da diesbezüglich Betrug weiterhin vorteilhaft ist. Es ergibt sich also wie für die anderen Konten-basierten Protokolle, dass ihnen aufgrund ihrer mangelnden Robustheit und größeren Komplexität das pessimistische Austauschprotokoll vorzuziehen ist.

2.3.2 Dritte zur Konfliktlösung

Bei den Austauschprotokollen zur Konfliktvermeidung wird der vertrauenswürdige Dritte in jeder Transaktion einbezogen. In der Literatur sind daher Protokolle vorgeschlagen worden, die nur in den Transaktionen auf den Dritten zurückgreifen, in denen es zum Betrug gekommen ist. Diese Protokolle sind in der Hinsicht *optimistisch*, dass der Dritte erst beim Auftreten eines Konflikts zwischen den Transaktionspartner aktiv wird und den *Konflikt* im Nachhinein *löst*.

Im Folgenden werden die Ansätze zu optimistischen Austauschprotokollen vorgestellt. Sie unterscheiden sich darin, welchen Grad der Fairness sie zu erreichen imstande sind.

Optimistische Austauschprotokolle für starke Fairness. Unter *starker Fairness* versteht man Fairness im engeren Sinne, nämlich dass am Ende der Transaktion beide oder keine der Einheiten von der Aktion des jeweiligen Transaktionspartners profitieren. Ein optimistisches Austauschprotokoll, das starke Fairness erreicht, wird in [Aso98] vorgeschlagen. Dazu geht der Ansatz davon aus, dass das Ergebnis einer Aktion als eine Zeichenkette i darstellbar ist. Abbildung 2.3 zeigt das Protokoll, dessen Schritte im Folgenden erklärt werden:

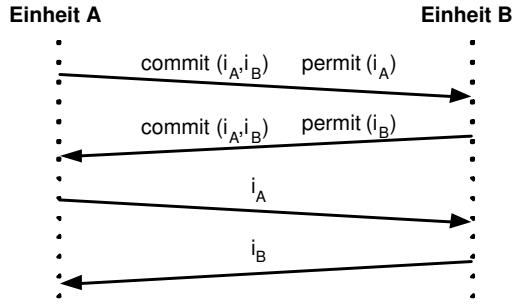

Abbildung 2.3: Optimistisches Austauschprotokoll für starke Fairness

1. Einheit A legt sich Einheit B gegenüber fest (engl.: commit), dass es in der Transaktion um den Austausch der Zeichenketten i_A und i_B geht. Dazu gibt A die Spezifikation der auszutauschenden i_A und i_B an. Hinzu fügt A eine Erlaubnis (engl.: permit) für B, die Zeichenkette i_A im Fall eines Konflikts durch den Dritten rekonstruieren zu lassen. Technisch gesehen wird dies ermöglicht, indem A sein i_A verschlüsselt an B weitergibt, so dass nur der Dritte zum Entschlüsseln in der Lage ist.

2. Analog zum ersten Schritt. Diesmal übergibt B seine Festlegung und Erlaubnis an A.

3. Einheit A händigt i_A an B aus.

4. Einheit B gibt sein i_B an A weiter.

Der Dritte wird nur dann in die Transaktion einbezogen, wenn es im Laufe des Protokolls zu Betrugsverhalten von einem der beiden Transaktionspartner kommt. In diesem Fall übergibt[6] der Dritte den Einheiten die auszutauschenden Zeichenketten i_A und i_B. Dafür verlangt der Dritte jedoch zuvor sowohl die Erlaubnisse für i_A und i_B als auch die Festlegungen beider Einheiten. Nach dem zweiten Schritt besitzen also beide Einheiten die Voraussetzungen, um die Zeichenkette ihres Gegenübers vom Dritten einzufordern. Interessant ist also lediglich der Fall, dass Einheit B den zweiten Schritt auslässt. Um dann dennoch an i_A zu gelangen, muss B dem Dritten seine Erlaubnis für i_B offenlegen. Dadurch gelangt auch Einheit A in den Besitz des erwünschten i_B. Wir erhalten also, dass das Protokoll starke Fairness garantiert.

Allerdings wird ein Problem offenbar, wenn wir den vierten Schritt des Protokolls untersuchen. Einheit B besitzt zu diesem Zeitpunkt schon i_A und hat somit keinen Anreiz, den Aufwand für das abschließende Übertragen von i_B zu tragen. Dadurch erhalten wir wie bei den Konten-basierten Protokollen, dass zumindest der letzte Schritt des Protokolls nicht ausgeführt wird. Da es also im letzten Schritt immer zu Betrug kommt, muss der Dritte in jeder Transaktion eingreifen. Dies bedeutet, dass das optimistische Protokoll keinen Vorteil gegenüber dem pessimistischen Protokoll bietet.

Eine Übersicht der Varianten zum Austauschprotokoll wird in [PSW98] gegeben. Sie nutzen spezielle Rahmenbedingungen des Systems (zum Beispiel die zeitlich garantierte Übertragung von Nachrichten) aus, um den Nachrichten- oder Zeitaufwand des Protokolls zu vermindern. Diese Varianten beseitigen jedoch nicht das Grundproblem, dass im letzten Protokollschritt immer betrogen wird.

[6] Den Schritt, aus einer Erlaubnis die ursprüngliche Zeichenkette wiederherzustellen, wird in der Literatur *generieren* genannt [Aso98].

Die bisherige Betrachtung ging davon aus, dass das Ergebnis einer Aktion eine Zeichenkette ist. Diese Annahme ist jedoch nicht erfüllt, wenn der Nutzen einer Aktion aus einem Seiteneffekt entsteht. Im Campus-Szenario ist das zum Beispiel dann der Fall, wenn ein Dienst zum Drucken von Dokumenten ausgeführt wird. Wegen dieser Annahme ist das optimistische Austauschprotokoll für starke Fairness nur beschränkt einsetzbar.

Optimistische Austauschprotokolle für schwache Fairness. Um mit allgemeinen Aktionen umgehen zu können, wird in existierenden Ansätzen der Begriff der Fairness abgeschwächt [Aso98]. Wir sprechen von *schwacher Fairness*, wenn Konflikte nicht durch den Dritten selbst sondern durch die Umgebung des Informationssystems gelöst werden können. Dazu werden im Laufe einer Transaktion *Beweismittel* über das Verhalten des Transaktionspartners gesammelt. Da Transaktionen im Informationssystem stattfinden, können nur *nicht-abstreitbare Marken* (engl.: non-repudiable token) solche Beweismittel darstellen. Wird eine Einheit betrogen, so kann sie mit ihren Beweismitteln bei einer externen Instanz (wie zum Beispiel einem Gericht) die Ausführung der Aktion, die vom Transaktionspartner versprochen worden ist, einfordern. Dadurch ändert sich die Rolle des Dritten: Er ist lediglich dafür da, die Transaktionspartner mit den entscheidenden Beweismitteln zu versorgen. Dazu muss der Dritte wahrnehmen können, ob die Aktionen der Transaktionspartner korrekt ausgeführt worden sind. Wir sprechen daher davon, dass für den Dritten das Verhalten der Transaktionspartner *nachvollziehbar* sein muss.

Ein Austauschprotokoll, das auf diese Idee baut, wird in [Aso98] vorgeschlagen[7]. Seine Schritte sind in Abbildung 2.4 dargestellt und werden im Folgenden erklärt:

1. Einheit A legt sich fest, welche Aktionen im Laufe der Transaktion ausgeführt werden.
2. Selbiges tut Einheit B.
3. Einheit A führt ihre Aktion aus und übermittelt ihr Ergebnis an B.
4. Auch Einheit B führt ihre Aktion aus. Außerdem stellt sie eine nicht-abstreitbare Quittung darüber aus, dass A's Aktion korrekt ausgeführt worden ist.
5. Einheit A bestätigt ihrerseits in einer Quittung, dass Einheit B ihre Aktion ausgeführt hat.

Das Protokoll unterscheidet sich nur in wenigen Punkten vom Protokoll für starke Fairness: **(1)** In den ersten beiden Schritten werden keine Erlaubnisse ausgestellt. Dies würde nämlich erfordern, dass es sich bei den Aktionen weiterhin um die Übergabe von Zeichenketten handelt. **(2)** Als Teil des vierten und fünften Schritts werden Quittungen ausgetauscht. Sie dienen als Beweismittel, mit denen sich die Einheiten nach der Transaktion gegen Anklagen verteidigen können.

Die Behandlung von Konflikten gestaltet sich wie folgt: Jede Einheit kann den Dritten durch das Vorzeigen beider Festlegungen dazu veranlassen, sich mit der Transaktion zu befassen. Sei A die Einheit, die den Dritten auf diese Art anruft. Der Dritte verlangt daraufhin von Einheit A, dass sie ihre Aktion ausführt und ihm eine an Einheit B gerichtete Quittung übergibt. Erfüllt A diese Forderungen, so wendet sich der Dritte an Einheit B und verlangt selbiges von ihr. Wenn B nicht antwortet, stellt der Dritte eine *Bestätigung* (engl.: affidavit) an Einheit A aus, dass B

[7] Die Darstellung des Protokolls in [Aso98] geht vom Austausch von Zeichenketten aus. In einem solchen Fall ergibt sich jedoch kein Vorteil im Vergleich zum optimistischen Protokoll für starke Fairness. Die weitere Besprechung des Protokolls für schwache Fairness geht daher von allgemeinen Aktionen aus.

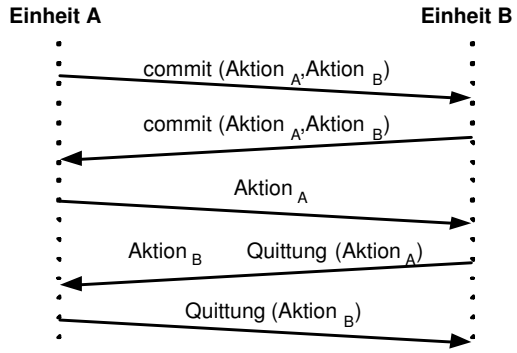

Abbildung 2.4: Optimistisches Austauschprotokoll für schwache Fairness

betrogen hat. Mit diesem Beweismittel kann Einheit A eine Anklage gegen B bei der externen Instanz einreichen und so das Ausführen von B's Aktion erzwingen.

Die Schwächen dieses Ansatzes liegen auf der Hand. Einerseits ist es sehr fragwürdig, ob der Dritte das Verhalten der Transaktionspartner beobachten kann. Dies ist aber die Voraussetzung dafür, dass Bestätigungen ausgestellt werden können. Andererseits wird in der Literatur nicht darauf eingegangen, wie die externe Instanz eine betrügende Einheit bestraft[8].

2.3.3 Dritte zur Bewertung und Verbreitung von Reputation

Eine andere Klasse von Ansätzen weist dem vertrauenswürdigen Dritten eine grundsätzlich andere Rolle zu: Er ist nicht dazu da, Konflikte im Vorfeld zu vermeiden oder nach ihrem Auftreten zu lösen. Hingegen bewertet und verbreitet der Dritte die *Reputation* der Einheiten, so dass er vor betrügenden Transaktionspartnern warnen kann.

In diesem Abschnitt werden die Ansätze besprochen, die von dieser Idee Gebrauch machen. Dazu gehen wir entlang der folgenden Fragestellungen vor:

- Wie ist die Architektur und Funktionsweise dieser reputationsbasierten Ansätze?
- Wie fassen die Einheiten Berichte über ihre Transaktionserfahrungen ab?
- Wie wird aus diesen Transaktionsberichten die Reputation der Einheiten abgeleitet?

Architektur und Funktionsweise. Abbildung 2.5 stellt die Architektur der reputationsbasierten Ansätze dar. Anhand der Architektur lässt sich die Funktionsweise der Ansätze aus der Sicht der Einheiten und des Dritten erklären:

- *Sicht einer Einheit:* Steht eine Transaktion mit einer anderen Einheit bevor, so wird beim Dritten die aktuelle *Reputation* dieses potentiellen Transaktionspartners nachgefragt. Nur wenn die Reputation des Partners hinreichend gut ist, zeigt sich die Einheit bereit, mit ihm in eine Transaktion zu gehen. Dies liegt daran, dass eine schlechte Reputation als ein Hinweis auf ein erhöhtes Betrugsrisiko aufgefasst wird. In dieser Hinsicht wird der Transaktionspartner entsprechend seiner Reputation *ausgewählt*. Nach der Transaktion stellt die

[8] Immerhin wird in [Aso98] ein Regelsystem vorgeschlagen, das die Richtigkeit von Anklagen aus der Menge vorgelegter Beweismittel ableiten kann. Der Ansatz gibt jedoch keinen Anhaltspunkt, wie sich das Ergebnis der Überprüfung auf die Beteiligten auswirkt.

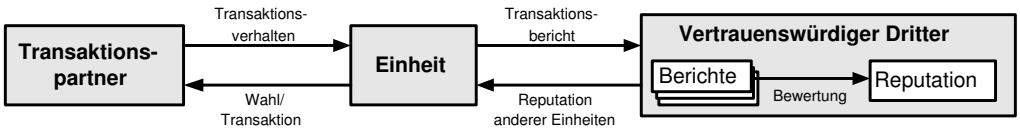

Abbildung 2.5: Architektur mit Bewertung und Verbreitung von Reputation durch Dritte

Einheit einen Bericht über das *Verhalten* des Transaktionspartners zusammen. In diesem steht, ob er sich kooperativ oder betrügend verhalten hat und in welchem *Kontext* sich die Transaktion bewegt hat [MHM02]. Ein wichtiger Bestandteil des Kontexts ist dabei, welche Art von Aktionen ausgeführt worden sind. Der Transaktionsbericht wird an den Dritten übergeben.

- *Sicht des Dritten:* Die Aufgabe des Dritten besteht darin, auf Anfrage die Reputation einzelner Einheiten mitzuteilen. Dazu sammelt er die Transaktionsberichte der Einheiten. Um zu einer *Bewertung* der Reputation der Einheiten zu gelangen, untersucht der Dritte den Inhalt der Transaktionsberichte. Zu diesem Zweck stellt er basierend auf diesen Berichten Überlegungen an, welches Verhalten die Einheiten in früheren Transaktionen wahrscheinlich gezeigt haben. Da die Berichte sich jeweils auf einen gewissen Kontext beziehen, ist auch die Bewertung der Reputation *kontextabhängig*.

In welcher Hinsicht wird durch diese Architektur das Betrugsverhalten der Einheiten eingeschränkt? Betrügt eine Einheit, so verfasst ihr Transaktionspartner einen entsprechenden Bericht an den Dritten. Als Folge davon ist eine Verschlechterung der eigenen Reputation zu erwarten. Dadurch wird es wiederum erschwert, Partner für zukünftige Transaktionen zu finden. Konkret bedeutet dies, dass die betrügende Einheit in einem verminderten Maße Zugang zu Informationen und Informationsdiensten anderer besitzt. Es zeigt sich also, dass durch die Verfügbarkeit von Reputation *Folgekosten* für Betrugsverhalten entstehen. Diese *Betrugskosten* (engl.: defection costs) geben einen Anreiz dafür, sich in Transaktionen kooperativ zu verhalten.

In der Literatur sind einige Ansätze bekannt, die auf dieser Architektur basieren. Sie werden *Reputationssysteme* genannt. Eine Übersicht dieser Ansätze wird in [Del03] gegeben. Der am weitesten verbreitete Ansatz ist das Bewertungssystem von *eBay* [eBa03]. Auf die beiden Schlüsselfragen, ob Transaktionsberichte wahrheitsgemäß verfasst werden und wie Reputation aus diesen Berichten abgeleitet wird, geben die Ansätze unterschiedliche Antworten. Daher werden diese beiden Punkte in den folgenden Paragraphen genauer untersucht.

Verfassen von Transaktionsberichten. Dem Dritten sind die Transaktionsberichte unabdingbar, um die Reputation der Einheiten bewerten zu können. Allerdings ist das Verfassen von Berichten unter zwei Aspekten problematisch. Diese werden im Folgenden vorgestellt.

Die Berichte der Einheiten müssen nicht notwendigerweise *wahrheitsgemäß* sein. Der Dritte kann einen Bericht nicht direkt auf seinen Wahrheitsgehalt hin überprüfen, da er das Verhalten der Transaktionspartner nicht beobachten kann. Dies ist insofern problematisch, als es für die Einheiten gute Gründe gibt, Berichte zu verfassen, die von dem tatsächlichen Transaktionsverhalten abweichen. Diese Gründe sind wie folgt:

- *Komplotte und Konkurrenten:* Eine Gruppe von Einheiten ist in der Lage, sich gegenseitig kooperatives Verhalten zu bescheinigen und sich dadurch hochzuloben. Eine solche Gruppe stellt ein *Komplott* (engl.: collusion) dar, weil sie gezielt den Dritten in die Irre führen will.

Auf der anderen Seite sind auch von denjenigen Einheiten keine wahrheitsgemäßen Berichte zu erwarten, die im Informationssystem direkte *Konkurrenten* bezüglich des Angebots oder der Nachfrage an Informationen haben. Jede Einheit diffamiert ihre Konkurrenten, um in mehr Transaktionen als Transaktionspartner ausgewählt zu werden.

- *Taktische Berichte:* In [Del03] wird darauf hingewiesen, dass in manchen Reputationssystemen die Berichte der Transaktionspartner eingesehen werden können. So hat zum Beispiel in eBay der Verkäufer Zugriff auf das Urteil des Käufers, bevor er selbst seinen Transaktionsbericht verfassen muss. In einem solchen Fall wird der Käufer aus Angst vor Vergeltung zu einem übermäßig positiven Transaktionsbericht verleitet. Ein solches *taktisches* Verhalten kann nur dann vermieden werden, wenn der Dritte die Transaktionsberichte vertraulich hält.

Der Wahrheitsgehalt von Berichten stellt nicht das einzige Problem dar: Das Verfassen eines Transaktionsberichts bringt für eine Einheit keinen Vorteil aber einen gewissen Aufwand mit sich [CCP98]. Daher ist fraglich, ob Transaktionsberichte überhaupt verfasst und dem Dritten *verfügbar* gemacht werden. Die existierenden Reputationssysteme begegnen diesem Problem dadurch, dass die Einheiten zum Verfassen von Berichten gezwungen werden [Del03] oder ihnen eine Belohnung dafür zugestanden wird [FKÖD04]. Außerdem ist es nicht in jedem Reputationssystem gewährleistet, dass Berichte nur über diejenigen Transaktionen verfasst werden können, die tatsächlich stattgefunden haben. So ist es zum Beispiel möglich, dass zwei Einheiten, die noch nie Transaktionspartner waren, sich gegenseitig Transaktionsberichte ausstellen. Es fehlt daher ein *Bezug* zwischen den Berichten und den Transaktionen, auf die sich die Berichte vermeintlich beziehen. Darunter leidet der Wahrheitsgehalt der Berichte.

Die Analyse zeigt also, dass der Dritte nur bedingt mit der Verfügbarkeit von Transaktionsberichten rechnen kann. Selbst wenn er einen Bericht erhält, so entspricht dieser möglicherweise nicht der Wahrheit.

Bewerten von Reputation. Um die Reputation von Einheiten auf eine aussagekräftige Art und Weise bewerten zu können, muss der vertrauenswürdige Dritte unterscheiden können, welche Transaktionsberichte wahrheitsgemäß sind und welche nicht. Als einziger Anhaltspunkt dient dem Dritten die *Plausibilität* des Berichts [Obr04a]. Die Techniken, die dazu in der Literatur vorgeschlagen werden, lassen sich entsprechend der Art der plausibilitätsbasierten Überlegungen in zwei Klassen einteilen[9]:

- *Endogene Bewertung:* Der *Inhalt* eines Berichts entscheidet darüber, ob er als wahrheitsgemäß eingestuft wird. Wird zum Beispiel ein negativer Bericht über eine Einheit abgegeben, die bisher immer als kooperativ befunden wurde, so erscheint der Bericht aufgrund seines Inhalts als nicht wahrheitsgemäß. Um solche *endogene* Bewertungen abgeben zu können, werden unterschiedliche Techniken eingesetzt. In [Del00] wird hierzu ein *Cluster-Filter Algorithmus* vorgeschlagen, der diffamierende Berichte identifiziert. Um auch hochlobende Berichte erkennen zu können, wird in [WJI04] eine *Beta-Verteilung* aus dem Inhalt der Berichte extrahiert und bewertet.

  Alle Techniken haben gemein, dass sie eine große Zahl von Transaktionsberichten benötigen, um zu zuverlässigen Ergebnissen zu kommen.

[9] Die Begriffsgebung der beiden Klassen entstammt [WJI04].

- *Exogene Bewertung:* Jeder Bericht wird anhand der *Vertrauenswürdigkeit* der berichtenden Einheit eingestuft. Da ein solches Verfahren vom Umfeld und nicht vom Inhalt des Berichts bestimmt wird, sprechen wir von einer *exogenen* Bewertung. Wie wird aber die Vertrauenswürdigkeit einer Einheit eingeschätzt? Eine triviale Möglichkeit besteht darin, sie der Reputation der Einheit gleichzusetzen. Allerdings muss dann auch die Richtigkeit ihrer Berichte in ihre Reputation einfließen [MHM02, ARH97, MT03].

  Hierbei ergibt sich ein grundlegendes Problem: Der Dritte kann das Transaktionsverhalten der Einheiten nicht direkt beobachten. Ob ein Bericht über eine Einheit wahrheitsgemäß ist, kann er daher nur anhand der Berichte anderer Einheiten einschätzen [WJI04]. Letztlich müssen also die Techniken der endogenen Bewertung zu Hilfe genommen werden.

Von ihrer Grundausrichtung her haben beide Klassen von Techniken eines gemeinsam. Sie gehen davon aus, dass die Mehrzahl der Berichte wahrheitsgemäß sind. Nur dann ist es möglich, aufgrund eines Vergleichs der Berichte die unwahren Berichte zu identifizieren. Allerdings ist diese Annahme nur dann erfüllt, wenn die meisten Einheiten auch tatsächlich Berichte verfassen und diese wahrheitsgemäß sind. Wir haben aber im vorigen Paragraphen einige Gründe dafür kennen gelernt, dass dies nicht der Fall ist. Damit ist die Effektivität der reputationsbasierten Ansätze zumindest fragwürdig.

2.3.4 Bewertung

Im Folgenden bewerten wir die in der Literatur vorgeschlagenen Ansätze aus der Sicht des Campus-Szenarios. Dabei gehen wir auf die Probleme ein, die bei der Verwendung vertrauenswürdiger Dritter entstehen. Das Fazit zeigt, dass die Selbstorganisation des Informationssystems und damit der Verzicht auf vertrauenswürdige Dritte wünschenswert ist.

Durchführbarkeit. Die Ansätze zur Konfliktvermeidung basieren auf pessimistischen Austauschprotokollen. Diese gehen davon aus, dass der Dritte Wissen über die Semantik der ausgetauschten Informationen besitzt. Erhält der Dritte zum Beispiel eine Information, so muss er überprüfen können, ob sie die Spezifikation erfüllt, die vor der Transaktion für die auszutauschenden Informationen vereinbart worden ist. Dies ist jedoch bei Informationen, die die teilnehmenden menschlichen Benutzer selbst erstellen, nicht gegeben. So ist es im Campus-Szenario nicht möglich, automatisiert zu überprüfen, ob zum Beispiel ein Dokument ein Vorlesungsmitschrieb ist oder nicht. Dadurch fehlt eine Voraussetzung für den Einsatz pessimistischer Austauschprotokolle.

Auch die Durchführbarkeit der Konten-basierten und optimistischen Austauschprotokolle ist stark eingeschränkt. Die Konten-basierten Protokolle sind nicht robust gegen Fehlverhalten und sind daher den pessimistischen Protokollen unterlegen. Die optimistischen Austauschprotokolle sind nur für ganz bestimmte Arten von Aktionen anwendbar, nämlich für Aktionen ohne Seiteneffekte und Aktionen, deren Ausführung vom Dritten beobachtet werden kann.

Ein weiteres Problem ergibt sich bei der Ausführung von Informationsdiensten. Die Austauschprotokolle beschränken sich darauf, dass die *Ergebnisse* der gegenseitig ausgeführten Informationsdienste fair ausgetauscht werden. Dies stellt zwar sicher, dass entweder keiner oder beide Transaktionspartner auf das Ergebnis der Informationsdienste zugreifen können. Jedoch kann diese Fairness nicht auf die *Ausführung* der Informationsdienste übertragen werden. Eine Einheit, die einen Informationsdienst erbracht hat, kann sich daher trotz fairem Austauschprotokoll nicht sicher sein, dass sie dafür eine Gegenleistung erhält. Dadurch wird der Nutzen der Austauschprotokolle stark eingeschränkt.

Aufgrund der beiden Mängel der Austauschprotokolle können die Ansätze zur Konfliktvermeidung und -lösung nur in Spezialfällen eingesetzt werden.

Verfügbarkeit des Dritten. Alle besprochenen Ansätze gehen davon aus, dass der vertrauenswürdige Dritte von den Teilnehmern erreicht werden kann. Der Grad der angenommenen Verfügbarkeit ist bei den Ansätzen unterschiedlich: Die Ansätzen zur Konfliktvermeidung beruhen darauf, dass der Dritte jederzeit verfügbar ist. Sonst können über ihn keine Transaktionen abgewickelt werden. Ähnliches gilt für die Ansätze der Konfliktlösung. Ihre Anforderungen an die Verfügbarkeit des Dritten sind allerdings nicht ganz so hoch, da der Dritte zur Lösung von Konflikten auch zeitlich leicht versetzt angerufen werden kann. Die reputationsbasierten Ansätze erfordern die Verfügbarkeit des Dritten, damit bei bevorstehenden Transaktionen die aktuelle Reputation des potentiellen Transaktionspartners abgefragt werden kann. Außerdem ist nach der Transaktion das Verhalten des Transaktionspartners zu bewerten und dem Dritten mitzuteilen.

Im Campus-Szenario ist allerdings die Verfügbarkeit eines vertrauenswürdigen Dritten stark eingeschränkt. So stellt sich die Situation typischerweise (anhand des Beispiels der Universität Karlsruhe) folgendermaßen dar: Auf dem Campus-Gelände hat die Universität ein WLAN-Netz aufgebaut, durch das Informationsgeräte sich ins Internet verbinden können [DUK00]. Auf den ersten Blick würde es also genügen, den vertrauenswürdigen Dritten auf einem Server im Internet verfügbar zu machen. Jedoch ist das WLAN-Netz in mehrfacher Hinsicht beschränkt:

- Das Campus-Gelände wird nur *in Teilen* von dem WLAN-Netz abgedeckt. Es gibt daher zahlreiche Stellen, an denen keine Internetverbindung besteht.
- Der Zugang zum WLAN-Netz ist auf wenige Geräte *beschränkt*. Kommt es zu einer örtlichen Konzentration der Geräte (zum Beispiel in Vorlesungsräumen), so lässt sich für viele Geräte keine Verbindung zum Internet aufbauen.
- Das WLAN-Netz ist *außerhalb* des Campus-Geländes *nicht verfügbar*. Informationsaustausch soll jedoch überall dort möglich sein, wo sich Informationsgeräte der Studenten treffen. Je nach Tages- und Jahreszeit befinden sich die Studenten mit ihren Informationsgeräten aber nicht so sehr auf dem Campus-Gelände selbst sondern in Wohngebieten, Parks, Verkehrsmitteln und dergleichen. An diesen Örtlichkeiten ist das WLAN-Netz aber nicht zugänglich. Damit kann der Dritte nicht über das Internet erreicht werden.

Wir erhalten also als Ergebnis, dass der vertrauenswürdige Dritte nicht in dem Maße verfügbar ist, wie es von den Ansätzen gefordert wird.

Aufwand seitens der Teilnehmer. Die infrastrukturbasierte Kommunikation mit einem vertrauenswürdigen Dritten ist auch unter einem weiteren Gesichtspunkt problematisch: Sie verursacht einen *höheren Energieverbrauch* als die direkte Kommunikation zwischen den Informationsgeräten [ZJPV02]. Dies liegt daran, dass der Energiebedarf für die drahtlose Kommunikation sich nach der Entfernung der Kommunikationspartner richtet. Damit ist die Kommunikation über einen entfernten Zugangspunkt zum Internet energetisch nicht effizient. Dies ist insofern nachteilig, da für Informationsgeräte die Kapazität ihrer Batterie die wertvollste (weil knappste) Ressource darstellt.

Ein weiteres Problem ergibt sich für die Teilnehmer des Informationssystems, wenn der Betreiber wie im Campus-Szenario nicht selbst für die *ubiquitäre Verfügbarkeit* des vertrauenswürdigen Dritten sorgen kann. In einem solchen Fall müssen ihre Informationsgeräte sich über kostenpflichtige Netze wie GPRS und UMTS [UMT03] mit dem Dritten verbinden. Die dabei anfallenden

Gebühren stellen jedoch den Mehrwert des Informationssystems in den Schatten. Außerdem ist ein negativer psychologischer Effekt auf die Studenten zu erwarten, da die Teilnahme am Informationssystem und der Informationsaustausch darin nicht mehr kostenlos ist.

Die Analyse zeigt also, dass die Teilnehmer des Informationssystems auf die Kommunikation mit einem vertrauenswürdigen Dritten gerne verzichten würden. Dadurch lässt sich nämlich der kritische Energieverbrauch einschränken und die Benutzung des Informationssystems bleibt kostenlos.

Aufwand seitens des Betreibers. Der Einsatz eines vertrauenswürdigen Dritten ist auch für den Betreiber des Informationssystems mit Kosten verbunden. Für den Betrieb des Dritten reicht es nämlich nicht aus, einen Server bereitzustellen, der mit dem Internet verbunden ist und auf dem die Software des Dritten installiert ist. Darüber hinaus muss folgendes gewährleistet sein:

- *Ausfallsicherheit:* Auf den Server muss jederzeit aus dem Internet zugegriffen werden können. Ansonsten lässt sich im gesamten Informationssystem keine einzige Transaktion durchführen. Diese Forderung stellt vor allem die Wartungsarbeiten am Server vor einer großen Herausforderung.

- *Skalierbarkeit:* Die Ressourcen des Servers müssen sich nach der Teilnehmerzahl und dem Transaktionsvolumen des Informationssystems richten. Im Campus-Szenario ist es wahrscheinlich, dass ein einfacher Server nicht ausreicht, um der anfallenden Last standzuhalten. Stattdessen wird eine Server-Farm benötigt, auf die der Dritte verteilt wird. Das Aufstellen und Warten einer solchen Server-Farm ist jedoch insbesondere unter dem Gesichtspunkt der Ausfallsicherheit sehr kostenintensiv.

- *Robustheit:* Die Teilnehmer des Informationssystems bauen darauf, dass der Dritte tatsächlich vertrauenswürdig ist. Dies ist jedoch nicht mehr gegeben, wenn die Server des Betreibers über das Internet angegriffen und unter die Kontrolle des Angreifers gebracht werden. Unter den obigen Forderungen der Ausfallsicherheit und Skalierbarkeit ist es besonders aufwändig, eine solche Robustheit zu erreichen.

Für den Betreiber fallen weitere Kosten an, wenn er eine hinreichende Verfügbarkeit des Dritten *gewährleisten* will. Dazu müsste das WLAN-Netz auf dem Campus-Gelände bezüglich seiner Abdeckung und Leistungsfähigkeit ausgebaut werden. Außerdem ist die nötige Erweiterung des Netzes auf Gebiete auch außerhalb des Campus besonders kostenintensiv und insofern problematisch, als die Universität dort nicht über die Grundstücke und Einrichtungen frei verfügen kann.

Es zeigt sich also, dass auch aus der Sicht des Betreibers des Informationssystems der Einsatz eines vertrauenswürdigen Dritten *nicht wünschenswert* ist.

Fazit. Die Bewertung der Ansätze mit vertrauenswürdigen Dritten ist in Tabelle 2.1 zusammengefasst:

- Ansätze zur *Konfliktvermeidung* führen zu einem hohen Aufwand für Teilnehmer und Betreiber des Informationssystems. Außerdem sind diese Ansätze nicht effektiv, da der Dritte nicht wie gefordert ständig verfügbar ist und die ausgetauschten Informationen nicht vom Dritten überprüfbar sind.

Tabelle 2.1: Bewertung der Ansätze zu vertrauenswürdigen Dritten

Eigenschaften der Ansätze	**Konflikt-vermeidung**	**Konflikt-lösung**	**Bewertung & Verbreitung von Reputation**
Durchführbarkeit	gering	gering	mittel
Benötigte Verfügbarkeit des Dritten	sehr hoch	hoch	mittel
Aufwand der Teilnehmer	sehr groß	mittel	groß
Aufwand des Betreibers	groß	mittel	groß

- Ein ähnliches Bild ergibt sich für die Ansätze zur *Konfliktlösung*. Lediglich der Aufwand für die Teilnehmer und den Betreiber des Informationssystems verringert sich etwas, da der Dritte nur im Konfliktfall kontaktiert wird.

- Ansätze, die auf die *Bewertung und Verbreitung von Reputation* durch ein Dritten basieren, sind am ehesten einsetzbar. Dies liegt daran, dass sie keine ständige Erreichbarkeit des Dritten erfordern und die Teilnehmer selbst die ausgetauschten Informationen überprüfen. Allerdings stellt es ein Problem dar, dass eine Einheit nur dann von der Reputation eines potentiellen Transaktionspartners erfahren kann, wenn der vertrauenswürdige Dritte erreichbar ist. Außerdem ergibt sich ein erhöhter Aufwand für die Teilnehmer und den Betreiber des Informationssystems, da die Bewertung und Berücksichtigung von Reputation weitaus komplexer ist als die Techniken, die von den Ansätzen zur Konfliktvermeidung und -lösung angewendet werden.

Es zeigt sich also, dass sich keiner der Ansätze für das Campus-Szenario eignet. Daher ist der Verzicht auf vertrauenswürdige Dritte wünschenswert. Im nächsten Abschnitt werden daher Ansätze besprochen, die auf eine selbstorganisierende Kontrolle der Teilnehmer untereinander zielen.

2.4 Systeme mit verteilter Vertrauensbildung

Um Betrugsverhalten einzudämmen, muss das Informationssystem Einfluss auf die teilnehmenden Einheiten ausüben. Die Ansätze aus dem letzten Abschnitt setzen zu diesem Zwecke vertrauenswürdige Dritten ein. Allerdings hat sich in jenem Abschnitt auch gezeigt, dass der Verzicht auf jegliche zentrale Instanz für den Betrieb des Informationssystems wünschenswert ist.

In der Literatur sind daher Ansätze vorgeschlagen worden, die diesem Problem begegnen. Sie zielen darauf, dass die Einheiten sich *gegenseitig* kontrollieren. Eine solche *selbstorganisierende soziale Kontrolle* erfordert, dass jede Einheit sich ihren Glauben über die Vertrauenswürdigkeit anderer Einheiten bildet. Die vorgeschlagenen Ansätze haben daher gemein, dass sie Mechanismen zur *verteilten Vertrauensbildung* ausarbeiten.

Abschnitt 2.4.1 bespricht die Grundsätze der verteilten Vertrauensbildung. Die Probleme, die bei ihr entstehen und bewältigt werden müssen, sind in Abschnitt 2.4.2 zusammengefasst. Es folgt eine Besprechung der Ansätze, die sich mit der Lösung dieser Probleme befassen. Die Darstellung in Abschnitt 2.4.3 und 2.4.4 unterscheidet dabei zwischen *qualitativen* und *quantitativen* Ansätzen. Abschließend werden die Ansätze in Abschnitt 2.4.5 dahingehend bewertet, ob sie den Anforderungen an die verteilte Vertrauensbildung gerecht werden.

2.4.1 Einführung

Der Ausgangspunkt für die verteilte Vertrauensbildung sind die Ansätze, die einen vertrauenswürdigen Dritten zur Bewertung und Verbreitung von Reputation verwenden. Durch den Verzicht auf den Einsatz eines Dritten muss jede Einheit in der Lage sein, das Verhalten anderer Einheiten *eigenständig* zu bewerten und zukünftige Transaktionspartner einzuschätzen. Anstatt der global gültigen Reputation der Ansätze mit Dritten ergibt sich dadurch eine spontane Herausbildung von Vertrauensbeziehungen zwischen den Einheiten. Eine solche *verteilte Vertrauensbildung* besitzt zwei entscheidende Vorteile über die reputationsbasierten Ansätze aus Abschnitt 2.3.3:

- *Zentrale Instanz überflüssig:* Die verteilte Vertrauensbildung kommt ohne einen vertrauenswürdigen Dritten aus. Dadurch tritt das Problem der mangelnden Verfügbarkeit des Dritten gar nicht erst auf. Außerdem ergibt sich eine nachhaltige Verringerung des Aufwands seitens der Teilnehmer, da sie sich nicht mehr mit einem entfernten Dritten abstimmen müssen. Besonders vorteilhaft stellt sich die Situation für den Betreiber des Informationssystems dar. Er muss weder für den Betrieb des Dritten noch für seine Verfügbarkeit sorgen. Außerdem die zentrale Speicherung von Verhaltensinformation, die von einem datenschutzrechtlichen Standpunkt aus als kritisch aufgefasst werden muss [MO04].

- *Basis für die Bewertung von Berichten:* Bei den reputationsbasierten Ansätzen ist die Methode der exogenen Bewertung von Erfahrungsberichten besonders viel versprechend, da sie mit einer geringeren Zahl von Informationsquellen auskommt. Allerdings wird in [WJI04] darauf hingewiesen, dass zur Bewertung der Berichte die Vertrauenswürdigkeit des Berichtenden eingeschätzt werden muss. In den reputationsbasierten Ansätzen ist dies problematisch, da die Bewertung ein Dritter durchführt, der keine eigenen Transaktionserfahrungen mit dem Berichtenden haben kann. Dieses Problem ist bei der verteilten Vertrauensbildung nicht gegeben. Dort führt jede Einheit selbst die Bewertung durch. Dazu kann sie ihre eigenen Transaktionserfahrungen heranziehen. Damit ist eine bessere Grundlage für die Bewertung von Berichten gegeben.

Der Rest dieses Abschnitts beschäftigt sich mit der allgemeinen Funktionsweise der verteilten Vertrauensbildung. Die *lokale Vertrauensbildung* besteht darin, dass jede Einheit eine Glaubensbildung basierend auf ihre eigenen Transaktionserfahrungen durchführt. Um eine mit den reputationsbasierten Ansätzen vergleichbare Menge von Erfahrungen für die Bewertung anderer heranziehen zu können, wird zusätzlich ein Empfehlungssystem benötigt. Die Empfehlungen und die darauf basierende Glaubensbildung erweitern die lokale Vertrauensbildung zu einer *verteilte Vertrauensbildung*.

Lokale Vertrauensbildung. Abbildung 2.6 zeigt die Funktionsweise der lokalen Vertrauensbildung aus der Sicht einer Einheit. Die Vertrauensbildung ist in einem Kreislauf organisiert, dessen Schritte im Folgenden besprochen werden.

Im Laufe einer *Transaktion* nimmt eine Einheit das Verhalten ihres Transaktionspartners wahr. Dieser verhält sich entweder *kooperativ*, indem er sich wie verabredet in der Transaktion verhält, oder *betrügend*. Das Wissen über das Transaktionsverhalten anderer legt die Einheit bei sich lokal ab und verwaltet sie, um zu einem späteren Zeitpunkt auf die eigenen Erfahrungen zugreifen zu können.

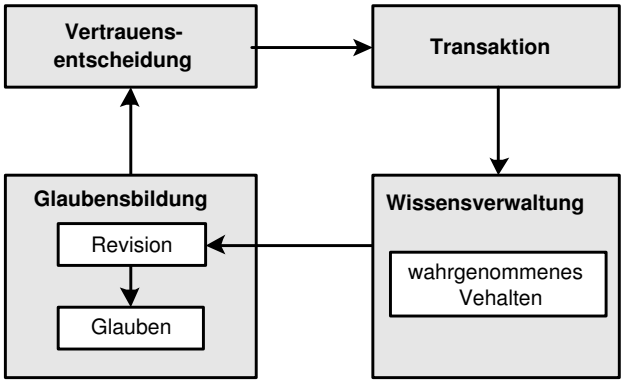

Abbildung 2.6: Der Kreislauf der lokalen Vertrauensbildung

Dies wird dazu benutzt, einen *Glauben* über wahrscheinliches zukünftiges Verhalten des Transaktionspartners zu *bilden*. Zu diesem Zweck wird der bisherige Glauben über ihn entsprechend des gezeigten Transaktionsverhaltens *revidiert*.

Dieser Glauben spielt eine entscheidende Rolle, welche Einheiten als zukünftige Transaktionspartner akzeptiert werden. Woran liegt das? Begibt sich eine Einheit in eine Transaktion, so liefert sie sich der Ungewissheit darüber aus, ob sie von ihrem Transaktionspartner betrogen wird. Die Teilnahme an einer Transaktion setzt daher Vertrauen in den Transaktionspartner voraus [SY01]. In dieser Hinsicht stellt die Entscheidung über die Teilnahme eine *Vertrauensentscheidung* dar. Je nach ihrem Glauben entscheidet sich eine Einheit dafür oder dagegen, dem Transaktionspartner zu vertrauen.

In der Literatur bezieht sich der Begriff *Vertrauen* fälschlicherweise auch oft auf den *Glauben* der Einheiten [BCL+04]. Davon wird in dieser Arbeit bewusst abgewichen. Vertrauen ist nur in dem Moment zugegen, in dem eine Einheit sich einem Risiko aussetzt, das von einer anderen Einheit maßgeblich beeinflusst wird [Mar94]. Dies ist nur bei der Entscheidung zur Teilnahme an einer Transaktion gegeben, da sich dadurch eine Einheit etwaigem Betrugsverhalten des Transaktionspartners aussetzt. Die Glaubensrevision basierend auf Verhaltensinformation findet jedoch abseits eines jeden Risikos statt und ist daher nicht mit dem Begriff Vertrauen zu bezeichnen.

Verteilte Vertrauensbildung. Die Glaubensbildung wird auf eine aussagekräftigere Basis gestellt, wenn die Einheiten sich gegenseitig von ihren Erfahrungen berichten. Abbildung 2.7 stellt die dadurch entstehende verteilte Vertrauensbildung aus der Sicht einer Einheit dar. Der Unterschied zur lokalen Vertrauensbildung besteht darin, dass die Einheiten zusätzlich *Empfehlungen* austauschen. Dem in der Abbildung dargestellten oberen Zyklus der lokalen Vertrauensbildung wird dadurch ein unterer Zyklus hinzugefügt, der die Verteilung der Transaktionserfahrungen zum Ziel hat.

Bei einer Empfehlung handelt es sich um Folgendes: Eine Einheit (der *Empfehler*) macht eine *Aussage* über eine andere Einheit (der *Empfohlene*). Die Aussage einer Empfehlung muss nicht unbedingt positiv sein, wie es im gängigen Sprachgebrauch üblich ist. Die Einheit, an die die Empfehlung übergeben wird, wird *Adressat* genannt. Er benutzt die Empfehlung, um seinen Glauben zu revidieren.

Alle bestehenden Ansätze haben gemein, dass die Empfehlungen keine direkte Aussage über Transaktionserfahrungen machen. Hingegen *berichtet* der Empfehler über seinen *Glauben* bezüglich des Empfohlenen. Der Adressat führt diesen Glaubensbericht seiner Wissensverwal-

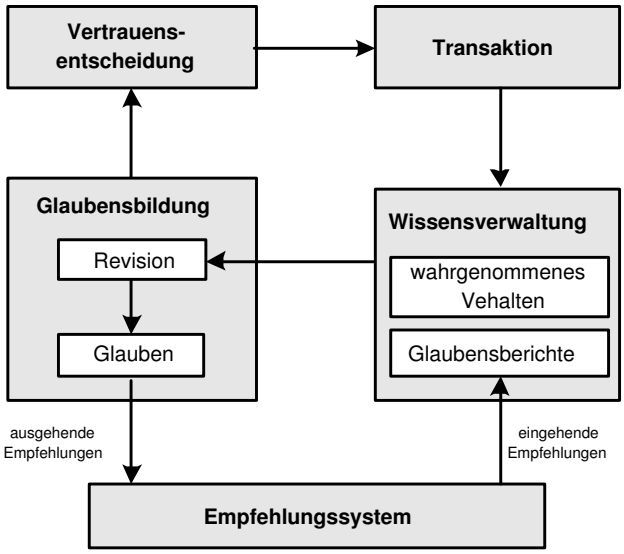

Abbildung 2.7: Der Kreislauf der verteilten Vertrauensbildung

tung zu. Jede Einheit verfügt daher über zwei Quellen für ihre Glaubensbildung, nämlich die erhaltenen Glaubensberichte und die eigene Wahrnehmung über das Verhalten der bisherigen Transaktionspartner.

2.4.2 Problembereiche und Anforderungen

Die verteilte Vertrauensbildung besitzt entscheidende Vorteile gegenüber den reputationsbasierten Ansätzen mit Dritten. Um dieses Potential umzusetzen, sind einige Probleme der verteilten Vertrauensbildung zu lösen. In diesem Abschnitt werden diese Probleme vorgestellt. Sie fallen in den Bereich der Glaubensbildung, des Empfehlungssystems und der Robustheit gegen Fehlverhalten. Aus den Problemen werden Anforderungen an die Ansätze zur verteilten Vertrauensbildung abgeleitet.

2.4.2.1 Glaubensbildung

Die Glaubensbildung bildet die Grundlage des Kreislaufs der lokalen Vertrauensbildung. Im Folgenden werden die Probleme beschrieben, die bei der Konzeption der Verfahren der Glaubensbildung berücksichtigt werden müssen.

Quantitative Glaubensbildung. Jede Einheit bildet sich ihren Glauben über die anderen Einheiten, um Vertrauensentscheidungen fällen zu können. Für eine Transaktion stehen prinzipiell zwei Entscheidungsmöglichkeiten zur Auswahl, nämlich die *Teilnahme* an der Transaktion und der *Verzicht* auf sie. Eine Einheit wird von diesen beiden Alternativen diejenige nehmen, die für sie einen größeren *erwarteten Residualnutzen*[10] verspricht. Da ein Verzicht weder zu Kosten noch zu Nutzen führt, erhalten wir als Kriterium für die Teilnahme an Transaktionen, dass der erwartete Residualnutzen für die Teilnahme *positiv* ist. Werden Vertrauensentscheidungen anhand dieses Kriteriums getroffen, so nennen wir sie *utilitaristisch*.

[10]Der Residualnutzen einer Alternative ist die Differenz zwischen dem Nutzen und den Kosten, die durch die Alternative verursacht werden.

Zur Berechnung des erwarteten Residualnutzens einer Transaktion ist die Wahrscheinlichkeit einzuschätzen, mit der der Transaktionspartner betrügen wird. Der Zweck der Glaubensbildung liegt eben hierin. Sie ist dafür zuständig, das zukünftige Verhalten Anderer auf eine *probabilistisch fundierte* Weise einzuschätzen. Eine Glaubensbildung, die diese Anforderung erfüllt, nennen wir *quantitativ*, da sie in der Lage ist, eine solche Wahrscheinlichkeit von Betrugsverhalten zu quantifizieren.

Nicht alle Verfahren der Glaubensbildung sind quantitativ. So ist auch eine *qualitative* Glaubensbildung denkbar, in der lediglich die Stärke des Glaubens ausgedrückt wird, die Wahrscheinlichkeit von Betrugsverhalten aber nicht quantifiziert werden kann. Folgendes Beispiel für eine solche Glaubensbildung findet sich in der Literatur weit verbreitet: Der Glauben darüber, ob eine Einheit sich zukünftig kooperativ verhalten wird, wird mit einem Wert (dem *Glaubenswert*[11]) im Intervall $[0, 1]$ oder $[-1, 1]$ angegeben. Je höher der Wert, desto wahrscheinlicher erscheint es, dass die Einheit sich zukünftig kooperativ verhalten wird. Es fehlt jedoch ein Zusammenhang zwischen diesem Wert und der Einschätzung einer Wahrscheinlichkeit. Er lässt sich auch durch Umrechnungen nicht herstellen, da der Wert bei der Glaubensrevision ohne Zuhilfenahme eines probabilistischen Modells angepasst wird.

Zusammenfassend fordern wir also, dass die Glaubensbildung quantitativ ist. Nur so lassen sich utilitaristische Vertrauensentscheidungen treffen.

Kontextabhängige Glaubensbildung. Der Kontext einer Transaktion bestimmt wesentlich, ob die Transaktionspartner zu kooperativem Verhalten bereit sind oder nicht. Dies lässt sich am wichtigsten Bestandteil des Kontexts, dem *Transaktionswert*, verdeutlichen:

- *Großer Transaktionswert:* Ist die Ausführung einer Aktion sehr aufwändig, so ergibt sich für die Einheit ein großer Anreiz, durch Unterlassen der Aktionsausführung zu betrügen.

- *Kleiner Transaktionswert:* Die Situation ändert sich, wenn die eigene Aktion kaum Aufwand verursacht. In einem solchen Fall ist es für eine Einheit eher attraktiv, sich kooperativ zu zeigen und damit Vertrauen bei ihrem Transaktionspartner aufzubauen. Dadurch ergeben sich nämlich unter Umständen weitere Gelegenheiten zu Transaktionen, deren Wert weitaus größer ist. Eine solche Strategie basiert also darauf, auf den geringen Betrugsvorteil bei Transaktionen mit kleinem Transaktionswert zu verzichten, um in den Genuss eines weitaus größeren Betrugsvorteils bei Transaktionen mit größerem Transaktionswert zu kommen.

Wir erkennen also, dass Betrugsverhalten kontextabhängig ist [MHM02]. Dies muss von der Glaubensbildung in zweierlei Hinsicht berücksichtigt werden: Einerseits ist beim Fällen von Vertrauensentscheidungen der Transaktionskontext einzubeziehen, um die Wahrscheinlichkeit von Betrugsverhalten einzuschätzen. Andererseits muss auch der Glauben über den Transaktionspartner in Abhängigkeit zum Kontext seines Verhaltens revidiert werden.

Berücksichtigung von Typinformation. Es gibt zwei *Typen*[12] von Einheiten, je nach dem welche Version der Systemsoftware auf ihrem Gerät installiert ist. Entweder ist dies die originale Systemsoftware oder eine manipulierte Version von ihr. Der Typ einer Einheit entscheidet also

[11] Dieser Wert wird in der Literatur häufig Vertrauenswert (engl.: trust value) genannt. Entsprechend des Arguments aus Abschnitt 2.4.1 ist die Bezeichnung Glaubenswert sinngemäßer.

[12] Dieser Typbegriff wird in Abschnitt 5.3.1 genauer definiert. An dieser Stelle wird nur so weit vorgegriffen, wie es für die Besprechung des Stands der Forschung nötig ist.

darüber, ob sie in der Lage ist zu betrügen (manipulierte Version) oder nicht (originale Version). Informationen über den Typ anderer Einheiten stellen daher einen wichtigen Anhaltspunkt über ihr zukünftiges Verhalten dar. Solche *Typinformationen* sind also in die Glaubensbildung einzubeziehen.

Es gibt zwei Möglichkeiten, wie eine Einheit zu Informationen über den Typ anderer Einheiten gelangt:

- *Typinformation aus dem Informationssystem:* Durch Beobachtung des Verhaltens anderer Einheiten kann eine Einheit zu Typinformationen gelangen. Als Voraussetzung dafür muss es gewisse Verhaltensmuster geben, die nur von Einheiten eines gewissen Typs gezeigt werden. Dies ist dann gegeben, wenn eine Einheit auf beabsichtigte Weise betrügt[13]. Daraus lässt sich nämlich ableiten, dass die Einheit eine manipulierte Version der Systemsoftware benutzt.

- *Typinformation aus dem realweltlichen Umfeld:* Unter gewissen Umständen besitzt ein menschlicher Benutzer verlässliche Informationen darüber, dass ein anderer Benutzer die originale Systemsoftware oder eine manipulierte Version auf seinem Gerät installiert hat. In einem solchen Fall macht er seiner Einheit die Information über den Typ der Einheiten dieser Benutzer zugänglich. Die Situationen, in denen der menschliche Benutzer an entsprechende Informationen gelangt, stellen sich wie folgt dar:

  - *Realweltliche Beziehungen:* Einige Benutzer des Informationssystems stehen in sozialen Beziehungen, die sich außerhalb des Informationssystems gebildet haben. Beispiele hierfür sind Beziehungen zwischen Verwandten oder engen Bekannten untereinander. Benutzer, die in solchen Beziehungen stehen, wissen unter Umständen darüber Bescheid, welche Version der Software von ihren Beziehungspartnern verwendet werden. Dies ist insbesondere dann der Fall, wenn sich Benutzer gegenseitig physischen Zugriff auf ihre Informationsgeräte (zum Beispiel zu Wartungszwecken) gestatten.
  - *Realweltliche Rollen:* In der Umgebung des Informationssystems spielen menschliche Benutzer unterschiedliche Rollen. So befinden sich im Campus-Szenario zum Beispiel die Studenten und die Angestellten der Universität in grundsätzlich verschiedenen Rollen. In der Literatur wird darauf hingewiesen, dass für manche dieser Rollen abgeleitet werden kann, welche Version der Systemsoftware auf dem Gerät eines Benutzers installiert ist [KSGM03]. Beim Campus-Szenario gilt dies für Angestellte, die Informationsgeräte der Universität ohne Administratorrechte benutzen. Bei ihnen kann davon ausgegangen werden, dass sie die originale Version der Systemsoftware verwenden.

Die Berücksichtigung von Typinformationen für die Glaubensbildung ist keineswegs trivial. Dies liegt daran, dass keine direkte Beziehung zwischen dem Typ und dem Verhalten einer Einheit besteht. Einerseits können sich auch Einheiten mit originaler Systemsoftware unbeabsichtigterweise fehlverhalten. Andererseits sind Einheiten mit einer manipulierten Version in der Lage, sich kooperativ zu verhalten. Daher kann die Wahrscheinlichkeit von Betrugsverhalten einer Einheit, deren Typ bekannt ist, nicht direkt auf 100% (bei originaler Systemsoftware) oder 0% (bei einer manipulierten Version davon) eingeschätzt werden. Die Verfahren der Glaubensbildung müssen dennoch eine Möglichkeit bieten, Typinformationen einzubeziehen.

[13] Dass für bestimmtes Betrugsverhalten eine Absicht nachgewiesen werden kann, wird in der Besprechung von inkonsistentem Verhalten in Abschnitt 7.3.1 dargelegt.

2.4.2.2 Empfehlungssystem

Die verteilte Vertrauensbildung verlangt danach, dass die Einheiten ihre Erfahrungen untereinander in Empfehlungen zugänglich machen. Das Empfehlungssystem gibt vor, wie Empfehlungen auszustellen und zu bewerten sind. Die Problempunkte, die dabei vermieden werden müssen, werden im Folgenden behandelt. Wie schon bei den reputationsbasierten Ansätzen mit Dritten in Abschnitt 2.3.3 gezeigt, nehmen hierbei die Verfügbarkeit und der Wahrheitsbezug der Empfehlungen eine zentrale Stellung ein.

Verfügbarkeit von Empfehlungen. Eine Einheit wird aus einer Reihe von Gründen davon abgebracht, ihre Transaktionserfahrungen auch anderen Einheiten zugänglich machen. Zum einen verursacht das Ausstellen und Verbreiten einer Empfehlung einen Aufwand, den der Empfehler selbst tragen muss. Zum anderen geht der Empfehler das Risiko ein, dass der Adressat die Empfehlung als unwahr einschätzt und dadurch das Vertrauen zu ihm verliert. Dies ist insbesondere dann der Fall, wenn der Empfehler eine Erfahrung gemacht hat, die den Erfahrungen anderer Einheiten widerspricht. Als Folge davon werden genau diejenigen Erfahrungen nicht weiter gegeben, die für andere Einheiten einen Neuigkeitswert enthalten würden.

Es ist nicht nur so, dass das Ausstellen von Empfehlungen eine Reihe von Nachteilen mit sich bringt. Darüber hinaus ergibt für den Empfehler auch kein Vorteil. Ein solcher Vorteil wird jedoch benötigt, um die Nachteile für das Empfehlen in den Schatten zu stellen und so dennoch für die Verfügbarkeit von Empfehlungen zu sorgen. Ein Empfehlungssystem muss daher sicherstellen, dass das Ausstellen von Empfehlungen einen Vorteil mit sich bringt.

Wahrheitsbezug von Empfehlungen. Wenn sich eine Einheit zum Empfehlen entscheidet, ist keinesfalls gewährleistet, dass der Inhalt der Empfehlung ihre eigenen Erfahrungen widerspiegelt und die Empfehlung damit einen *Wahrheitsbezug* besitzt. Speziell für die bestehenden Empfehlungssysteme bedeutet das, dass die in einer Empfehlung enthaltenen Glaubensberichte nicht dem Glauben des Empfehlers entsprechen müssen.

Eine Einheit hat eine Reihe von Beweggründen, Empfehlungen auszustellen, die nicht wahrheitsgemäß sind. In Abschnitt 2.3.3 haben wir bereits entsprechende Gründe für die Erfahrungsberichte dargestellt, die von reputationsbasierten Ansätzen mit Dritten verwendet werden: Einheiten können ein Komplott bilden, indem sie sich gegenseitig kooperatives Verhalten bescheinigen. Weiterhin diffamieren sich Einheiten gegenseitig, die untereinander in einer Konkurrenzsituation stehen. Zudem gibt es eine Reihe von subtileren Möglichkeiten, sich durch taktisches Empfehlen einen Vorteil im Informationssystem zu verschaffen.

In einem Empfehlungssystem, wie es von der verteilten Vertrauensbildung eingesetzt wird, kommt ein weiteres Problem hinzu. Im Gegensatz zu den reputationsbasierten Ansätzen mit Dritten besitzt keine Einheit eine globale Sicht darüber, welche Einheiten welche Empfehlungen abgegeben haben. Es ist einer Einheit daher möglich, an unterschiedliche Einheiten sich gegenseitig widersprechende (also *inkonsistente)* Empfehlungen auszustellen, ohne dass dies erkennbar ist. Zum Beispiel kann Einheit A an Einheit C Einheit B als stets kooperative Einheit empfehlen, während sie Einheit D gegenüber Einheit B als Betrüger darstellt. Die Situation entspricht also dem Problem der *byzantinischen Generäle* [LSP82]. Damit ergeben sich weitergehende Möglichkeiten für taktisches Empfehlungsverhalten.

Ein weiteres Problem entsteht, wenn das Empfehlungssystem dadurch die Verfügbarkeit von Empfehlungen sicherstellt, dass das Empfehlen an sich belohnt wird [JF03, FKÖD04]. In einem solchen Fall ergibt sich für die Einheiten ein Anreiz zum Empfehlen, auch wenn über den Ge-

genstand der Empfehlung keinerlei Erfahrungen vorliegen. Dadurch entbehren Empfehlungen im Vorhinein jeglichen Wahrheitsbezuges.

Es erscheint also, dass die Verfügbarkeit und der Wahrheitsbezug von Empfehlungen im Konflikt stehen. Das Ausstellen von Empfehlungen bietet nämlich nur dann einen Vorteil, wenn durch Komplotte, Konkurrenzsituationen oder taktischen Überlegungen empfohlen wird. In einem solchen Fall besitzen Empfehlungen jedoch keinen Wahrheitsbezug. Es ist die Aufgabe des Empfehlungssystems, eben diesen Konflikt zu lösen.

Fundierte Glaubensrevision. Das Empfehlungssystem setzt Empfehlungen ein, um eine weitere Informationsquelle für die Glaubensbildung zu erschließen. Eine Einheit, die eine Empfehlung erhält, soll also basierend auf der Empfehlung ihren Glauben revidieren.

Zu diesem Zweck muss das Empfehlungssystem ein Verfahren bereitstellen, mit dem die Glaubensrevision durchgeführt wird. Dieses Verfahren hat die Einschränkungen bezüglich der Verfügbarkeit und des Wahrheitsbezuges von Empfehlungen zu berücksichtigen. Trotz dieser Schwierigkeit muss die Glaubensrevision auf eine probabilistisch fundierte Weise durchgeführt werden, um eine quantitative Glaubensbildung zu gewährleisten.

2.4.2.3 Nachgewiesene Robustheit gegen Fehlverhalten

Im Zuge von Transaktionen können sich Einheiten fehlverhalten, in dem sie betrügen. Der Sinn der verteilten Vertrauensbildung besteht eben darin, dieses Fehlverhalten einzudämmen. Dabei ist aber darauf zu achten, dass durch den Einsatz der verteilten Vertrauensbildung nicht neue Möglichkeiten zu Fehlverhalten entstehen. Ist Fehlverhalten gegenüber der Vertrauensbildung dennoch möglich, so ist es von der Vertrauensbildung selbst einzudämmen[14]. Wir müssen daher fordern, dass die verteilte Vertrauensbildung *abgeschlossen* gegenüber Fehlverhalten ist [Obr04a].

Beim Entwurf der Glaubensbildung und des Empfehlungssystems sind *Überlegungen* darüber anzustellen, ob und warum gewisse Arten von Fehlverhalten durch den Entwurf vermieden werden. Hier zeigt sich in den Beiträgen in der Literatur zur verteilten Vertrauensbildung eine große Schwachstelle. Eine Untersuchung dieser Beiträge in [Sab03] ergibt nämlich, dass weniger als 30% von ihnen überhaupt Überlegungen über mögliches Fehlverhalten anstellen. Der Rest der Beiträge geht davon aus, dass die einzige Art von Fehlverhalten betrügendes Verhalten in Transaktionen ist.

Aber auch wenn Überlegungen über mögliches Fehlverhalten angestellt werden, ist die Robustheit eines Ansatzes nicht ausreichend belegt. Dies liegt daran, dass in solchen Überlegungen nicht alle Arten von Fehlverhalten antizipiert werden können. Es bedarf daher eines *Nachweises* der Robustheit eines jeden Ansatzes. Dieser ist entweder *analytisch* oder *simulativ* zu führen.

Der analytische Nachweis benutzt die Methoden der Spieltheorie. Das entscheidende Kriterium ist hierbei die *Anreizkompatibilität* [JF03]. Ein Verhaltensschritt ist für eine Einheit anreizkompatibel, wenn die Einheit nur an ihren eigenen Vorteil denkend freiwillig diesen Schritt ausführt. Im analytischen Nachweis ist also zu zeigen, dass alle Teile der Glaubensbildung und des Empfehlungsverhaltens anreizkompatibel sind.

Alternativ dazu lässt sich der Nachweis aus simulativ führen. Hierfür wird allerdings eine Methodik benötigt, mit der alle Arten von zu erwartendem Fehlverhalten systematisch herausge-

[14] Fehlverhalten gegenüber der Vertrauensbildung könnte zwar durch ein weiteres System zur Vertrauensbildung eingedämmt werden. Damit würde sich aber die Frage stellen, wie Fehlverhalten gegenüber diesem zweiten System eingeschränkt werden kann. Wir erhalten also ein rekursives Argument.

funden werden können. In der anschließenden Simulation wird der Entwurf der Glaubensbildung und des Empfehlungssystems auf die Robustheit gegenüber solchem Fehlverhalten überprüft.

2.4.3 Qualitative Ansätze

Die überwiegende Zahl der bestehenden Ansätze zur verteilten Vertrauensbildung bauen auf einer qualitativen Glaubensbildung auf. Solche Ansätze werden im Folgenden qualitativ genannt. Sie können zumindest zwei Anforderungen aus Abschnitt 2.4.2 nicht erfüllen: Zum einen ist die Glaubensbildung nicht wie gefordert quantitativ. Zum anderen kann die Berücksichtigung von Empfehlungen nicht zu einer fundierten Glaubensrevision führen, eben weil die Glaubensbildung auf keinem probabilistischen Modell basiert.

In diesem Abschnitt untersuchen wir, ob die bestehenden qualitativen Ansätze die weiteren Anforderungen an die verteilte Vertrauensbildung erfüllen. Es wird dabei nur auf solche Ansätze eingegangen, deren Funktionsweise in den entsprechenden Veröffentlichungen klar definiert ist. Arbeiten wie [KR03], die sich lediglich mit der Architektur der Vertrauensbildung befassen, werden also nicht berücksichtigt.

Die qualitativen Ansätze lassen sich anhand zweier Kriterien unterscheiden:

- *Anwendungsbereich:* Manche Ansätze entwerfen Verfahren der verteilten Vertrauensbildung, die für *spezielle* Anwendungsbereiche (wie zum Beispiel dem Routing) ausgelegt sind. Im Gegensatz dazu ist es aber auch möglich, die verteilte Vertrauensbildung als Mechanismus zu entwerfen, der für einen beliebigen Anwendungsbereich Verwendung finden kann. Das Modell der verteilten Vertrauensbildung, wie es in Abschnitt 2.4.1 vorgestellt wurde, geht von einer solchen *allgemein* einsetzbaren Vertrauensbildung aus.

- *Glaubensdomäne:* Die qualitativen Ansätze haben gemein, dass die Stärke des Glaubens in einem Glaubenswert Ausdruck findet, der keine probabilistische Interpretation erlaubt. Die Menge der möglichen Glaubenswerte nennen wir *Glaubensdomäne*. Einige Ansätze besitzen eine *enumerierbare* Glaubensdomäne, die nur bestimmte Glaubenswerte (wie zum Beispiel kooperativ oder betrügend) zulässt. Hingegen setzen andere Ansätze eine *kontinuierliche* Glaubensdomäne ein, indem sie Glaubenswerte aus dem Intervall $[-1, 1]$ oder dem Intervall $[0, 1]$ vorsehen.

Im Folgenden wird zunächst auf die Ansätze mit einem speziellen Anwendungsbereich eingegangen. Anschließend kommt es zur Untersuchung der Ansätze, die allgemein einsetzbar sind. Dabei wird zwischen Ansätzen mit enumerierbarer und kontinuierlicher Glaubensdomäne unterschieden.

Ansätze mit einem speziellen Anwendungsbereich. Es gibt zwei Anwendungsbereiche, für die spezielle Verfahren der verteilten Vertrauensbildung entworfen worden sind. Zum einen stellt sich beim *Routing* in Ad-hoc Netzen die Frage, wann ein Gerät an welches andere Gerät Nachrichten oder Teile von Nachrichten weiterleiten sollte [RT99]. Zum anderen müssen auch in *Overlay-Netzen* [ABKM01] Nachrichten weitergeleitet werden, so dass sich dort eine analoge Frage stellt. Glauben spielt in diesen Anwendungsbereichen insofern eine Rolle, als der Weiterleiter einer Nachricht davon überzeugt sein muss, dass **(1)** der Absender der Nachricht ihn für seine Bemühungen belohnt und **(2)** das Gerät, an das die Nachricht weitergeleitet wird, auch selbst die Nachricht weiterleiten wird.

Beim Routing und in Overlay-Netzen unterscheidet sich Kontext verschiedener Transaktionen nur marginal. Es ist daher nicht überraschend, dass die Ansätze hierfür auf eine Kontextabhängigkeit der Glaubensbildung verzichten.

Watchdog/Pathrater [MGLB00] war der erste Ansatz, der sich mit der Vertrauensbildung beim Routing auseinander setzte. Jede Einheit besitzt eine Komponente (der so genannte *Watchdog*), der das Verhalten anderer Einheiten beobachtet und als entweder kooperativ oder betrügend bewertet. Der Ansatz benutzt daher eine binäre Glaubensdomäne. Die Einheiten tauschen zwar ihre Glaubenswerte untereinander aus. Jedoch wird weder ein Anreiz zum Ausstellen von Empfehlungen noch zum wahrheitsgemäßen Empfehlen bereitgestellt. Trotz des simulativen Evaluation des Ansatzes kann kein Nachweis seiner Robustheit gegenüber Fehlverhalten erbracht werden, da die Möglichkeit taktischen Fehlverhaltens nicht berücksichtigt wird.

Core [MM02b, MM02a] baut den Ansatz von *Watchdog/Pathrater* aus, indem die Glaubensdomäne auf das Intervall $[-1, 1]$ verfeinert wird. Zudem werden nur noch positive Empfehlungen erlaubt, um Diffamierungen zu vermeiden. Dadurch wird jedoch weder die Verfügbarkeit noch der Wahrheitsbezug von Empfehlungen sichergestellt.

Eine andere Richtung wird von *Friends & Foes* [MR03] eingeschlagen. Als Erweiterung zu *Watchdog/Pathrater* wird hierbei den Einheiten erlaubt, gewisse Einheiten als Feinde (engl.: foes) zu kennzeichnen, deren Nachrichten nicht weitergeleitet werden. Eine Einheit wird nur dann als betrügend bewertet, wenn sie ihren Freunden (engl.: friends) gegenüber das Weiterleiten verweigert. Der Ansatz zielt also darauf, dass die Einheiten den Grad ihrer Beteiligung am Netz bestimmen können, ohne als betrügend aufgefasst zu werden. An den Schwächen der Glaubensbildung und des Empfehlungssystems von *Watchdog/Pathrater* ändert diese Erweiterung allerdings nichts.

Moreton et al. [MT03] stellt einen Ansatz zur Vertrauensbildung im Overlay-Netz *Kademlia* vor. Das Empfehlungssystem sieht einen gegenseitigen Austausch von Empfehlungen vor. Eine Einheit, die auf das Ausstellen von Empfehlungen verzichtet, erhält damit auch keinerlei Empfehlungen von anderen Einheiten. Damit wird ein Anreiz zum Ausstellen von Empfehlungen hergestellt. Allerdings lösen die Autoren dadurch nicht den Konflikt zwischen der Verfügbarkeit und dem Wahrheitsbezug von Empfehlungen. Unter dem Zwang, Empfehlungen auszustellen, wird nämlich eine Einheit genau so empfehlen, wie ihr gegenüber empfohlen worden ist. Sie wird dadurch als glaubwürdiger Empfehler aufgefasst, ohne dass ihre Empfehlungen einen Wahrheitsbezug haben.

Böhm et al. [BB04b, BB05] schlägt einen weiteren Ansatz für Overlay-Netze vor. Der Glaubenswert bezüglich einer Einheit wird durch die Zahl der positiven Erfahrungen mit ihr gleichgesetzt. Nur wenn der Glaubenswert einen bestimmten Schwellwert überschreitet, kommt es zu einer Transaktion mit der Einheit. Dadurch wird ein Verfahren benötigt, durch das auch Einheiten, über die nicht genügend positive Erfahrungen vorliegen (etwa weil sie Neulinge sind), in Transaktionen teilnehmen können. Die Autoren schlagen hierfür das Lösen *kryptographischer Puzzle* [ANL00] (so genannte *proof of work*) als Substitut für Vertrauenswürdigkeit vor. Konkret bedeutet dies, dass einer Einheit die Funktionalität des Overlay-Netz zur Verfügung gestellt wird, wenn **(1)** sie sich zuvor im Overlay-Netz hinreichend kooperativ verhalten hat oder **(2)** sie auf Anfrage ein Puzzle löst. Dieses Verfahren geht also davon aus, dass auch unkooperativen Einheiten gegenüber Dienste erbracht werden, solange diese den Aufwand des Puzzle-Lösens auf sich nehmen. Diese Annahme ist allerdings nicht tragbar. Aus der Sicht der anderen Einheiten gibt es nämlich keinen Anreiz, Dienste für die unkooperative Einheiten zu erbringen, da diese von keinem Nutzen für das Overlay-Netz sind. Insbesondere werden daher Anfragen für das Puzzle-Lösen niemals gestellt. Damit wird den Verfahren dieses Ansatzes die Grundlage entzogen.

Allgemeine Ansätze mit enumerierbarer Glaubensdomäne. Von den Ansätzen, die an keinen speziellen Anwendungsbereich gebunden sind, schlagen einige eine enumerierbare Glaubensdomäne vor. Diese Ansätze werden im Folgenden besprochen.

Der Ansatz von Castelfranchi et al. [CCP98, CP02] unterscheidet zwischen zwei Klassen von Einheiten. *Normative* Einheiten verhalten sich in Transaktionen immer kooperativ, während *strategische* Einheiten sich zum Betrug entscheiden, wenn dies Vorteile bietet. Diese Klassifizierung steht in direktem Zusammenhang zur Typisierung aus Abschnitt 2.4.2, die an der Version der benutzten Systemsoftware orientiert ist. Allerdings erkennen die Autoren diesen Zusammenhang nicht, sondern benutzen diese Klassifizierung der Einheiten lediglich für den Entwurf einer binären Glaubensdomäne. Dabei wird eine Einheit solange als normativ angesehen, bis ein Betrug von ihr beobachtet wird. Sodann erscheint sie als strategisch. Eine solche Glaubensbildung geht davon aus, dass normative Einheiten sich nicht unbeabsichtigterweise fehlverhalten. Entsprechend des Systemmodells aus Abschnitt 1.2.2 ist diese Annahme aber nicht haltbar. Der Vorteil der binären Glaubensdomäne besteht darin, dass Typinformationen direkt berücksichtigt werden können. Ein weiteres Kennzeichen des Ansatzes ist, dass das vorgeschlagene Empfehlungsmodell weder für die Verfügbarkeit noch für den Wahrheitsbezug von Empfehlungen sorgt.

Papaioannou et al. [PS05] führt eine andere binäre Glaubensdomäne ein. Eine Einheit steht entweder unter Strafe oder sie ist ungestraft. Wie von dieser Formulierung angedeutet, handelt es sich bei diesen beiden Glaubenswerten also um eine globale Eigenschaft. Zu diesem Zweck wird eine koordinierende zentrale Instanz benötigt. Diese sammelt die Berichte über die Transaktionserfahrungen der Einheiten. Eine Einheit wird für eine gewisse Zeitdauer unter Strafe gestellt, wenn sie keinen oder einen vom Transaktionspartner abweichenden Erfahrungsbericht ausstellt oder wenn sie eine Transaktion mit einer unter Strafe stehenden Einheit eingeht. Die zentrale Instanz ist insbesondere dafür zuständig, dass die beiden Transaktionspartner nicht gegenseitig ihre Erfahrungsberichte einsehen können. Nur dann kann die Verfügbarkeit und der Wahrheitsbezug von Empfehlungen hergestellt werden. Die Autoren behaupten zwar, dass die globale Eigenschaft des Gestraft-Seins durch ein verteiltes Protokoll hergestellt werden kann. Allerdings zeigen sie nicht, wie in einem solchen Fall der Zugang zu Erfahrungsberichten gewährt beziehungsweise verweigert werden kann. Insgesamt muss dieser Ansatz daher als in einem selbstorganisierenden Informationssystem nicht durchführbar eingestuft werden.

Allgemeine Ansätze mit kontinuierlicher Glaubensdomäne. Viel versprechender sind diejenigen allgemeinen Ansätze, die eine kontinuierliche Glaubensdomäne verwenden. Damit lässt sich nämlich die Stärke des Glaubens genauer festhalten. Im Folgenden werden solche Ansätze untersucht. Ebenso wie die bisher besprochenen Ansätze handelt es sich hierbei weiterhin um qualitative Ansätze der verteilten Vertrauensbildung. Obwohl die Stärke des Glaubens festgehalten wird, lassen sich also die Glaubenswerte selbst nicht probabilistisch interpretieren.

EigenTrust [KSGM03] benutzt das Intervall $[0, 1]$ als Glaubensdomäne. Die Glaubenswerte einer Einheit werden so normiert, dass sie sich zu 1 addieren. Der Sinn einer solchen Normierung liegt darin, die Berechnung eines global gültiger Glaubenswerte zu erleichtern. Dazu wird allerdings eine zentrale Instanz benötigt, da die von den Autoren vorgeschlagenen Verfahren zur verteilten Berechnung nicht robust gegen Fehlverhalten sind[15]. Außerdem wird weder für die Verfügbarkeit noch für den Wahrheitsbezug von Empfehlungen gesorgt.

[15]Dies gilt auch für ihren Vorschlag, dass mehrere Einheiten für diese globale Berechnung zuständig sind. Einerseits kann nicht gewährleistet werden, dass diesen Einheiten genau dieselben Glaubensberichte als Quelle der Berechnung zugestellt werden. Andererseits ist es unklar, wie eine Einheit, deren Berechnungen von denen der anderen Einheiten abweicht, bestraft werden kann.

Marti et al. [MGM04] verzichtet auf die Berechnung globaler Glaubenswerte und kommt dadurch im Gegensatz zu *EigenTrust* ohne zentrale Instanz aus. Allerdings gibt es weiterhin keinen Anreiz zum wahrheitsgemäßen Empfehlen.

Liu et al. [LI04] schlägt als Glaubensdomäne das Intervall $[-1, 1]$ vor. Empfehlungen finden entsprechend der Glaubwürdigkeit des Empfehlers Eingang in die Glaubensrevision. Die Plausibilität des Inhalts der Empfehlungen entscheidet darüber, wie glaubwürdig der Empfehler erscheint. Damit ergibt sich wie bei Moreton et al. [MT03], dass Einheiten genau so empfehlen, wie ihnen gegenüber empfohlen worden ist. Das Empfehlungssystem kann damit nicht die Verfügbarkeit und den Wahrheitsbezug von Empfehlungen sicherstellen. Eine Besonderheit der Glaubensbildung liegt darin, dass sie kontextabhängig ist. Zu diesem Zweck gehen die Autoren davon aus, dass die Menge von Kontexten in einem Baum angeordnet werden kann. Für jeden Kontext wird ein eigener Glaubenswert gehalten. Bei Erhalt von Verhaltensinformation kommt es bei all denjenigen Kontexten zur Glaubensrevision, die den Kontext des beobachteten Verhaltens verallgemeinern. Dies führt zu einer äußert aufwändigen Glaubensverwaltung. Die Autoren können die Robustheit des Ansatzes gegen Fehlverhalten nicht nachweisen, da ihre simulative Evaluation nur einfache Arten von Fehlverhalten berücksichtigt.

Das *Buddy-System* [FOKR04, FO04] teilt viele seiner Eigenschaften mit Liu et al. Dies gilt sowohl für die Glaubensbildung mitsamt ihres Kontextmodells als auch für das Empfehlungssystem. Eine Besonderheit des Buddy-Systems liegt darin, dass sich eine Einheit auch selbst empfehlen kann, indem sie die Einheiten, die ihr vertrauen, (so genannte *Buddies*) auflistet. Durch solche Vertrauensbeziehungen (so genannte *Buddy-Beziehungen*) wird die *soziale Struktur* im Informationssystem explizit gemacht [OFN04]. Außerdem besteht ein inhärenter Anreiz für die Bekanntgabe der eigenen Vertrauensbeziehungen, da das Ausstellen von Selbstempfehlungen für den Empfehler selbst vorteilhaft ist. Der Erweiterung des Empfehlungssystems um Selbstempfehlungen sorgt also für einen gewissen Grad an Verfügbarkeit und Wahrheitsbezug von Empfehlungen. Jedoch sieht das Buddy-System ein ad-hoc Verfahren für die Glaubensrevision vor, das den Möglichkeiten der Buddy-Beziehungen nicht gerecht wird.

2.4.4 Quantitative Ansätze

In diesem Abschnitt befassen wir uns mit denjenigen Ansätzen der verteilten Vertrauensbildung, deren Glaubensbildung quantitativ ist. Sie legen darauf Wert, dass das zukünftige Verhalten anderer Einheiten auf eine probabilistisch fundierte Weise eingeschätzt werden kann. Daher ist im Gegensatz zu den qualitativen Ansätzen der Glaube durch eine Wahrscheinlichkeit anzugeben. Hierfür wird ein probabilistisches Modell benötigt, das aus Informationen über vergangenes Verhalten zukünftiges Verhalten bezüglich seiner Wahrscheinlichkeit vorhersagt.

Im Folgenden werden die bestehenden quantitativen Ansätze besprochen. Sie unterscheiden sich hauptsächlich darin, welches probabilistischen Modell sie für die Glaubensbildung einsetzen.

Yu et al. Die *Dempster-Shafer* Theorie bildet die Grundlage für die Glaubensbildung in Yu et al. Diese Theorie befasst sich damit, wie aus Beobachtungen die Wahrscheinlichkeit für das Zutreffen verschiedener Hypothesen abgeleitet wird. Unter anderem kann eine Beobachtung mehrere oder keine Hypothese unterstützen.

Die erste Schwäche des Ansatzes ergibt sich daraus, dass als Hypothesen die Aussagen *Einheit X ist vertrauenswürdig* (kurz $\mathtt{T}$) und *Einheit X ist nicht vertrauenswürdig* (kurz $\neg\mathtt{T}$) gewählt werden. Damit werden die Belange der Glaubensbildung mit denen der Vertrauensentscheidungen vermischt. Außerdem werden die Beobachtungen von Transaktionsverhalten direkt mit jeweils

einer Hypothese identifiziert, nämlich kooperatives Verhalten mit T und betrügendes Verhalten mit ¬T. Dadurch wird von der Stärke der Dempster-Shafer Theorie, eine Beobachtung verschiedenen Hypothesen probabilistisch korrekt zuzuschreiben, nicht Gebrauch gemacht. Außerdem lassen sich Kontext- und Typinformationen nicht in das Glaubensmodell integrieren.

Die zweite Schwäche des Ansatzes liegt in seinem Empfehlungsmodell. Es verpasst nicht nur, für die Verfügbarkeit und den Wahrheitsbezug von Empfehlungen zu sorgen. Darüber hinaus sind die vorgeschlagenen Verfahren der Glaubensrevision mit zwei Nachteilen verbunden: **(1)** Der in einer Empfehlung enthaltene Glaubensbericht wird durch *Dempsters Kombinationsregel* mit dem eigenen Glauben kombiniert. Diese Regel nimmt die Glaubensberichte anderer Einheiten als gleichwertig zum eigenen Glauben an und berücksichtigt dadurch nicht, dass Glaubensberichte unwahr sein können. **(2)** Das Ergebnis der Kombination weist auch der Leerhypothese, dass weder T noch ¬T zutrifft, eine positive Wahrscheinlichkeit zu, die für die Glaubensbildung nicht interpretiert werden kann[16]. Dieses Problem entsteht aus der unsachgemäßen Anwendung der Dempster-Shafer Theorie bei der Definition der Hypothesen.

In der simulativen Evaluation wird nur eine Art von Fehlverhalten berücksichtigt. Daher bleibt der Ansatz den Nachweis seiner Robustheit gegenüber beliebigem Fehlverhalten schuldig.

Sen et al. Die in Sen et al. [SS02b] vorgeschlagene Glaubensbildung beruht auf einem einfachen Verfahren. Seien $(r_1, ..., r_n)$ die n Transaktionserfahrungen, die über eine Einheit vorliegen. Dabei bedeutet $r_i = 1$ (beziehungsweise $r_i = 0$), dass die Einheit sich kooperativ (beziehungsweise betrügend) verhalten hat. Damit wird die Wahrscheinlichkeit für kooperatives zukünftiges Verhalten durch $\sum_i r_i \cdot \alpha^i$ mit einem positiven Faktor $\alpha < 1$ berechnet. Durch diesen Faktor werden jüngere Erfahrungen stärker gewichtet. Information über Kontext und Typ werden von der Glaubensbildung nicht betrachtet.

Im Fokus des Ansatzes steht das Empfehlungssystem. Es geht davon aus, dass ein Großteil der Einheiten immer wahrheitsgemäße Empfehlungen ausstellt. Dadurch wird es einer Einheit möglich, durch das Einholen einer gewissen Zahl von Empfehlungen mit hoher Wahrscheinlichkeit sicherzugehen, dass der Empfohlene sich mit einer Wahrscheinlichkeit von mindestens 50% kooperativ verhalten hat. Die Probleme eines solchen Empfehlungssystems sind offensichtlich: **(1)** Der Ansatz setzt die Verfügbarkeit und den Wahrheitsbezug von Empfehlungen voraus, anstatt dafür durch geeignete Verfahren zu sorgen. **(2)** Die Wahrscheinlichkeit für kooperatives Verhalten des Transaktionspartners lässt sich nur als größer oder kleiner als 50% angeben. Was fehlt, ist also ein differenzierteres Modell für Empfehlungen und die Glaubensrevision.

Despotovic et al. Der Ansatz von Despotovic et al. [DA04] führt ein anderes Verfahren zur Glaubensbildung ein. Es geht davon aus, dass kooperatives und betrügendes Verhalten einer jeden Einheit *Bernoulli*-verteilt ist. Dies ist der Fall, wenn eine Einheit in jeder Transaktion mit einer gewissen intrinsischen Wahrscheinlichkeit p betrügt und diese Wahrscheinlichkeit über all ihre Transaktionen hinweg gleich bleibt. Die Aufgabe der Glaubensbildung liegt also darin, die für jede Einheit spezifische Wahrscheinlichkeit p zu ermitteln.

Das Empfehlungssystem des Ansatzes ist eng mit der Glaubensbildung verbunden. In Empfehlungen werden keine Glaubensberichte mitgeteilt. Hingegen sind darin die Transaktionserfahrungen mit dem Empfohlenen aufgeschlüsselt. Damit kennt der Adressat der Empfehlung auch die Transaktionserfahrungen des Empfehlers $(r_1^E, ..., r_m^E)$. Die Erfahrungsberichte werden zusammen

[16] Halten zum Beispiel beide Einheiten die Wahrscheinlichkeit von T und ¬T für jeweils 50%, so ergibt die Kombination, dass T und ¬T jeweils nur noch mit 25% wahrscheinlich sind.

mit den eigenen Erfahrungen dazu benutzt, die intrinsische Wahrscheinlichkeit p des Empfohlenen zu erhalten. Sie berechnet sich mit Hilfe eines Likelihood-Schätzers so, dass die eigenen und berichteten Transaktionserfahrungen so wahrscheinlich wie möglich erscheinen. Zu diesem Zweck ist die Glaubwürdigkeit eines jeden Empfehlers einzuschätzen.

Der Ansatz besitzt zwei Stärken. Einerseits ist die Glaubensbildung probabilistisch fundiert. Dies äußert sich unter anderem darin, dass die eigenen Transaktionserfahrungen so in die Glaubensbildung einfließen, als ob sie von einem vollständig glaubwürdigen Empfehler stammen. Andererseits ist das vorgeschlagene Verfahren zur Einschätzung der Glaubwürdigkeit von Empfehlern den zuvor vorgestellten plausibilitätsbasierten Ansätzen überlegen. Ein Empfehler erscheint nämlich glaubwürdig, wenn der Inhalt seiner Empfehlung in einer anschließenden Transaktion mit dem Empfohlenen sich bewahrheitet.

Der größte Schwachpunkt des Ansatzes ist seine Abhängigkeit von der Verfügbarkeit von wahrheitsgemäßen Empfehlungen. Zudem ist die Einschätzung der Glaubwürdigkeit eines Empfehlers nur dann möglich, wenn dieser zuvor schon einige Empfehlungen über eine Einheit ausgestellt hat, mit der eine Reihe von Transaktionserfahrungen vorliegen. Unter normalen Umständen lässt sich also die Glaubwürdigkeit eines Empfehlers nicht genau einschätzen. Dadurch wird die Präzision der Glaubensrevision stark beeinträchtigt.

Die simulative Evaluation des Ansatzes stellt die Prämisse, dass jede Einheit sich mit einer bestimmten Wahrscheinlichkeit fehlverhält, nicht in Frage. Damit werden einige Arten von Fehlverhalten nicht berücksichtigt. Es wird also kein Nachweis für die Robustheit des Ansatzes gegenüber Fehlverhalten erbracht.

Confidant. *Confidant* [BB02b, BB03, BB04a] benutzt von allen quantitativen Ansätzen das überzeugendste probabilistische Modell für die Glaubensbildung. Es geht wie in Despotovic et al. davon aus, dass jede Einheit mit einer gewissen Wahrscheinlichkeit in ihren Transaktionen betrügt. Diese Wahrscheinlichkeit wird für jede Einheit wie folgt ermittelt: Sei γ die Zahl der positiven und δ die Zahl der negativen Transaktionserfahrungen mit einer Einheit. Aufbauend auf den Arbeiten des *Beta Reputation System* [JI02] wird daraus abgeleitet, dass die Wahrscheinlichkeitsdichtefunktion (engl.: probability density function) für die Einheit einer Beta-Verteilung folgt mit $\beta(\gamma, \delta)$. Mit dieser Funktion lassen sich probabilistisch fundierte Aussagen über wahrscheinliches zukünftiges Verhalten einer Einheit treffen. Allerdings hat dieses probabilistische Modell zum Nachteil, dass Kontext- und Typinformation nicht einbezogen werden kann.

Eine entscheidende Schwäche von *Confidant* ist sein Empfehlungssystem. Eine erhaltene Empfehlung findet nur dann Eingang in die Glaubensrevision, wenn sie hinreichend plausibel erscheint. Wie bei Moreton et al. richtet sich die Plausibilität einer Empfehlung nach der Glaubwürdigkeit des Empfehlers und damit indirekt danach, wie sehr der Inhalt der Empfehlung dem eigenen Glauben entspricht. Damit ergibt sich auch hier, dass Einheiten genau so empfehlen, wie ihnen gegenüber empfohlen worden ist. Die Verfügbarkeit und der Wahrheitsgehalt von Empfehlungen wird also nicht gewährleistet. Zudem arbeitet die Glaubensrevision mit einem ad-hoc Verfahren, um die als glaubwürdig erscheinenden Empfehlungen in den eigenen Glauben einzubeziehen.

In der simulativen Evaluation von *Confidant* bleiben einige Arten von viel versprechendem Fehlverhalten unberücksichtigt. Daher bleibt der Ansatz den Nachweis seiner Robustheit gegenüber solchem Fehlverhalten schuldig.

2.4.5 Bewertung

Die verteilte Vertrauensbildung zielt darauf, dass sich die Einheiten des Informationssystems gegenseitig kontrollieren. Von der Effektivität einer solchen selbstorganisierenden sozialen Kontrolle hängt ab, ob Betrugsverhalten im Informationssystem eingedämmt werden kann. Die besprochenen Ansätze müssen sich also daran messen, ob sie die Anforderungen an die verteilten Vertrauensbildung zu erfüllen in der Lage sind. Eine zusammenfassende Bewertung wird in Tabelle 2.2 gezeigt. Im Folgenden wird darauf eingegangen, wie sich die Situation für die einzelnen Anforderungen darstellt.

Die *Glaubensbildung* erfolgt nur in wenigen Ansätzen *quantitativ*. Eine solche quantitative Glaubensbildung ist jedoch für das Treffen von utilitaristischen Vertrauensentscheidungen unabdingbar. Informationen über den *Kontext* werden nur von zwei qualitativen Ansätzen berücksichtigt. Zudem ist nur der Ansatz von Castelfranchi et al. durch seine binäre Glaubensdomäne in der Lage, *Typinformationen* zu berücksichtigen.

Die Schwäche der bestehenden Ansätze offenbart sich vor allem in den von ihnen vorgeschlagenen *Empfehlungssystemen*. Kein Ansatz erreicht eine Lösung des Konflikts zwischen der *Verfügbarkeit* und dem *Wahrheitsbezug* von Empfehlungen. Allein das *Buddy-System* schafft hierfür zum Teil Abhilfe, indem es das Empfehlungssystem um Selbstempfehlungen erweitert. Jedoch sind die dazu verwendeten Verfahren nicht ausgereift, so dass der Konflikt nicht gänzlich gelöst wird. Eine probabilistisch *fundierte Glaubensrevision* basierend auf Empfehlungen wird nur von Despotovic et al. vorgeschlagen. Allerdings geht der Ansatz durch seine Annahmen von zu günstigen Rahmenbedingungen für die Glaubensrevision aus.

Alle bestehenden Ansätze haben gemein, dass sie ihre *Robustheit* gegenüber Fehlverhalten nicht *nachweisen* können. Zwar unternehmen die Ansätze Versuche zur Erbringung eines solchen Nachweises. Diese scheitern aber aus unterschiedlichen Gründen. Der analytische Nachweis aus Papaioannou et al. basiert auf einer Reihe von unhaltbaren Annahmen. Die anderen Ansätze führen eine simulative Evaluation durch, um den Nachweis zu erbringen. Dabei werden allerdings nicht alle Arten von viel versprechendem Fehlverhalten berücksichtigt.

Keiner der bestehenden Ansätze erfüllt also die Anforderungen an die verteilte Vertrauensbildung. Damit können diese Ansätze nicht erreichen, dass Betrugsverhalten effektiv vermieden wird.

Auch durch Kombination der Ansätze lässt sich dieses negative Ergebnis nicht verändern. Es gibt nämlich einige Anforderungen, die von keinem der Ansätze bewältigt werden. Hinzu kommt, dass die Ansätze aufgrund ihrer unterschiedlichen Glaubensbildung nicht kombinierbar sind. So lassen sich zum Beispiel *Typinformationen* im Ansatz von Castelfranchi et al. nur deswegen berücksichtigen, weil seine Glaubensbildung qualitativ ist. Wir kommen daher zum Schluss, dass nur durch einen grundlegend andersartigen Ansatz die verteilte Vertrauensbildung zu einer effektiven Betrugsvermeidung führen kann.

2.5 Fazit

Im Campus-Szenario ist das Informationssystem vollständig auf die Geräte der Endbenutzer verteilt. In diesem Kapitel haben wir uns die Frage gestellt, inwieweit bestehende Forschungsansätze die Realisierung dieser Vision ermöglichen. Dabei ist ein besonderes Augenmerk darauf gelegt worden, wie der Zielkonflikt zwischen Selbstorganisation des Informationssystems, Autonomie der teilnehmenden Einheiten und Effektivität der Betrugsvermeidung angegangen wird.

Eine Reihe von Ansätzen geht von der Benutzung *manipulationssicherer Hardware* aus,

Tabelle 2.2: Bewertung bestehender Ansätze zur verteilten Vertrauensbildung

Anforderungen \ Ansätze		Ansätze für Routing und Overlay -Netze	Castelfranchi et al.	Papaioannou et al.	EigenTrust Liu et al.	Liu et al.	Buddy System	Yu et al.	Sen et al.	Despotovic et al.	Confidant
Glaubensbildung	quantitativ	-	-	-	-	-	-	+	+	+	+
	kontext - abhängig	-	-	-	-	+	+	-	-	-	-
	berücksichtigt Typinformation	-	+	-	-	-	-	-	-	-	-
Empfehlungssystem	Verfügbarkeit	-	-	-	-	-	o	-	-	-	-
	Wahrheits - bezug	-	-	-	-	-	o	-	-	-	-
	fundierte Glaubensrevision	-	-	-	-	-	-	-	-	o	-
Nachgewiesene Robustheit		-	-	-	-	-	-	-	-	-	-

Legende:
+ Anforderung erfüllt
o Anforderung zum Teil erfüllt
- Anforderung nicht erfüllt

um den menschlichen Benutzern die Kontrolle über ihre Gerät zu entziehen. Durch diese Einschränkung der Autonomie der teilnehmenden Einheiten wird von Vornherein verhindert, dass die Systemsoftware manipuliert wird. Damit lässt sich in der originalen Systemsoftware individuell irrationales Verhalten wie der Verzicht auf Betrugsverhalten durchsetzen. Es gibt zwei grundsätzliche Möglichkeiten zum Einsatz von manipulationssicherer Hardware, die aus unterschiedlichen Gründen nicht tauglich sind. Wird die Systemsoftware fest auf ein *spezielles Hardware-Modul* abgelegt, so entstehen dem Endbenutzer Kosten, die eine zu hohe Eintrittsbarriere für die Teilnahme am Informationssystem darstellen. *Mehrzweck-Hardware* würde zwar dieses Problem beseitigen, sie ist jedoch nicht verfügbar. Damit sind all jene Ansätze nicht tauglich, die einer Beschränkung der *Autonomie* der Teilnehmer bedürfen.

Eine andere Richtung wird von den Ansätzen verfolgt, die den Einsatz von *vertrauenswürdigen Dritten* vorschlagen. Als zentrale Instanz greift ein solcher Dritter regulierend in das Informationssystem ein, um Betrugsverhalten unvorteilhaft zu machen. Damit ist das Informationssystem nur noch in eingeschränkter Weise selbstorganisierend. Die Ansätze unterscheiden sich darin, welche Rolle sie dem Dritten zuweisen. Bei den Ansätzen der *Konfliktvermeidung* schaltet sich der Dritte in den Ablauf einer Transaktion ein, so dass es für Einheiten unmöglich ist zu betrügen. Hingegen wird in den Ansätzen der *Konfliktlösung* der Dritte nur dann einbezogen, wenn der Betrug eines Transaktionspartners dokumentiert oder rückgängig gemacht werden soll. Zu diesem Zweck wird das Verhalten der Transaktionspartner durch den Einsatz von nicht-abstreitbaren Marken für den Dritten *nachvollziehbar* gemacht. Die Ansätze zur Konfliktvermeidung und -lösung haben gemein, dass sie aufgrund der Annahmen der zugrunde liegenden Austauschprotokolle nur in Spezialfällen eingesetzt werden können. Einen anderen Weg gehen daher die *reputationsbasierten* Ansätze, bei denen der Dritte Informationen über das vergangene Verhalten der Einheiten sammelt, um betrügende Einheiten mit einer schlechten Reputation zu kennzeichnen. Allerdings werden dadurch die Nachteile für die Verwendung vertrauenswürdiger Dritter nicht beseitigt: Die Teilnehmer und

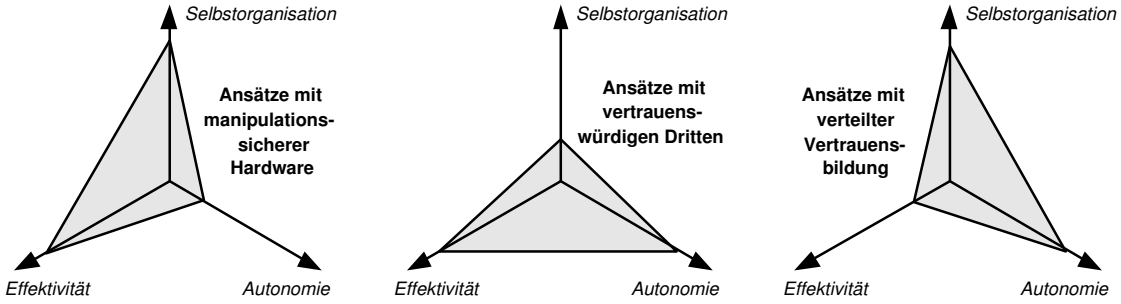

Abbildung 2.8: Unterschiedliche Schwachpunkte der bestehenden Forschungsansätze

der Betreiber des Informationssystems müssen einen enormen Aufwand tragen, damit der Dritte von den Einheiten durchgängig erreichbar ist. Damit sind all jene Ansätze nicht gangbar, die einer Einschränkung der *Selbstorganisation* des Informationssystems bedürfen.

Keinen Kompromiss bezüglich Autonomie und Selbstorganisation gehen daher die Ansätze zur *verteilten Vertrauensbildung* ein. Sie bewirken eine gegenseitige Kontrolle der Teilnehmer, indem jede Einheit sich ihren Glauben über die Vertrauenswürdigkeit Anderer bildet. Allerdings verpassen es die Ansätze, eine Reihe von Problemen zu bewältigen. Das Schlüsselproblem besteht darin, wie die *Erfahrungen Anderer* in die eigene Glaubensbildung einfließen können. In allen Ansätzen stellen Empfehlungen billiges Gerede dar und erlauben somit keine glaubwürdige Signalisierung eigener Erfahrungen. Die Glaubensbildung kann dadurch nur mit ad-hoc Verfahren durchgeführt werden, die auf Plausibilitätsüberlegungen basieren. Der Glaubensbildung mangelt es daher an notwendiger Präzision und an Robustheit gegenüber Fehlverhalten. Damit können die Ansätze der verteilten Vertrauensbildung nicht erreichen, dass Betrugsverhalten effektiv vermieden wird.

Abbildung 2.8 fasst zusammen, wie die Ansätze den Zielkonflikt zwischen Selbstorganisation des Informationssystems, Autonomie der teilnehmenden Einheiten und Effektivität der Betrugsvermeidung angehen. Die Randbedingungen der Autonomie und Selbstorganisation stellen sicher, dass der Aufwand für die Inbetriebnahme und des Betriebs des Informationssystems vertretbar ist. Darüber hinaus ermöglicht erst eine effektive Betrugsvermeidung die Existenzfähigkeit von Informationssystemen. Die bestehenden Ansätze erreichen jedoch jeweils nur zwei der drei Zielvorgaben. Der Zielkonflikt wird also von keinem der Ansätze gelöst. Dadurch lässt sich unsere Vision von Informationssystemen, die wie im Campus-Szenario vollständig auf die Geräte der Endbenutzer verteilt sind, nicht mit Hilfe bestehender Ansätze realisieren.

Kapitel 3

君子 主信 無友不好己者

"Der Edle muss lernen, Versprechen zu halten und die Freundschaft all derer abzuweisen, die nicht so wie er sind (indem sie Profit und Verlust als Ausgangspunkt ihres Handelns machen)."
(Gespräche und Aussprüche des Konfuzius, Auszug aus 1.8)

Ansatz dieser Arbeit

Die Vision von Informationssystemen, die wie im Campus-Szenario vollständig auf die Geräte der Endbenutzer verteilt sind, lässt sich mit bestehenden Ansätzen nicht realisieren. Dies liegt daran, dass kein Ansatz den Zielkonflikt zwischen Selbstorganisation des Informationssystems, Autonomie der teilnehmenden Einheiten und Effektivität der Betrugsvermeidung löst. Um Betrug effektiv zu vermeiden, geben Ansätze mit Dritten oder mit manipulationssicherer Hardware die Selbstorganisation des Informationssystems oder die Autonomie der teilnehmenden Einheiten auf. Ansätze der verteilten Vertrauensbildung garantieren zwar Selbstorganisation und Autonomie. Sie erreichen aber wegen grundsätzlicher Schwächen im Ansatz keine effektive Vermeidung von Betrugsverhalten.

Dieses Kapitel stellt vor, wie diese Arbeit dieses Problem beseitigt und dadurch die Validierung unserer These über die Existenzfähigkeit der betrachteten Informationssysteme ermöglicht. Dazu werden im Folgenden die Grundideen vorgestellt, auf die sich der *Entwurf* des Ansatzes stützt (Abschnitt 3.1). Der Ansatz geht von der verteilten Vertrauensbildung aus und erweitert sie um Schlüsselideen der Systeme mit manipulationssicherer Hardware und mit Dritten. Diese Erweiterung löst den Zielkonflikt, da sie eine effektive Betrugsvermeidung herstellt, ohne dass durch den Einsatz von Dritten oder von manipulationssicherer Hardware die Selbstorganisation oder Autonomie eingeschränkt werden muss. Anschließend wird eine Übersicht der Grundideen zur *Evaluation* des Ansatzes gegeben (Abschnitt 3.2). Sie ermöglichen Aussagen darüber, unter welchen Rahmenbedingungen unsere These validiert werden kann.

3.1 Grundideen des Entwurfs

Wie in Abschnitt 2.4 gezeigt, sind existierende Vorschläge zur verteilten Vertrauensbildung nicht in der Lage, Betrugsverhalten effektiv zu vermeiden. Dennoch bietet sich verteilte Vertrauensbildung als Ausgangspunkt des eigenen Ansatzes an, da sie *selbstorganisierende soziale Kontrolle* zwischen *autonomen* Einheiten ermöglicht. Es kommt also darauf an, Erweiterungen und Anpassungen zu finden, die diese soziale Kontrolle und damit die Betrugsvermeidung effektiv machen.

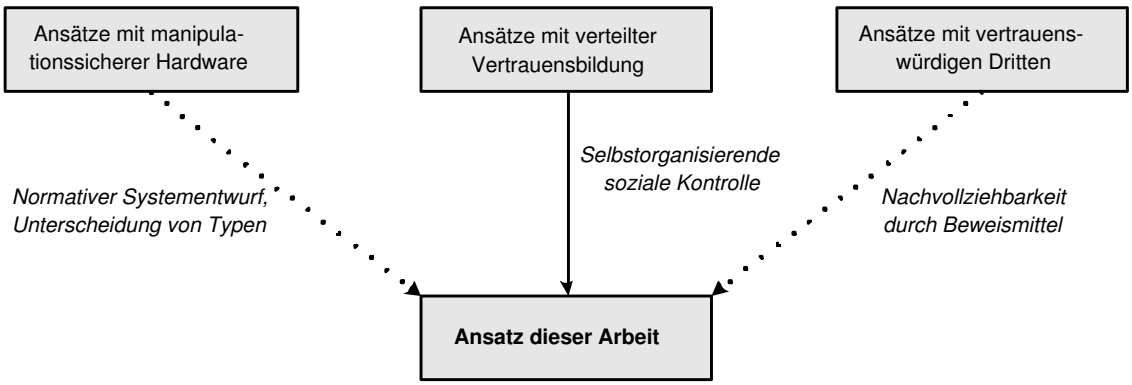

Abbildung 3.1: Einordnung des eigenen Ansatzes

In Abbildung 3.1 werden die Grundsätze für den Entwurf des eigenen Ansatzes eingeordnet. Er ist ein *quantitativer* Ansatz der verteilten Vertrauensbildung. Wie in den Ansätzen aus Abschnitt 2.4.4 lassen sich also Vertrauensentscheidungen unter Zuhilfenahme subjektiver Einschätzungen des potentiellen Transaktionspartners utilitaristisch treffen. Um die soziale Kontrolle effektiv zu machen, werden zusätzlich Schlüsselideen der Ansätze mit manipulationssicherer Hardware und mit Dritten in den eigenen Ansatz aufgenommen.

Ein Kennzeichen von *manipulationssicheren Systemen* ist, dass der Systementwurf auch Verhalten vorsehen darf, das individuell nicht rational ist. Das liegt daran, dass der menschliche Prinzipal seine Einheit nicht manipulieren kann, auch wenn ihr Verhalten nicht in seinem Sinne ist. Der Systementwurf besteht also aus dem Aufstellen von *Normen*, die die Verhaltensmöglichkeiten einschränken [CCP98]. Ausgehend auf der Beobachtung, dass auch ohne manipulationssicherer Hardware der Akt der Manipulation aufwändig ist, kann ein solcher *normativer Systementwurf*[1] auch für den eigenen Ansatz angewendet werden. Es ergeben sich weit reichende Folgen für das Modell der *Glaubensbildung*, das im Gegensatz zu der Verhaltensorientierung existierender Ansätze auf den *Typ* einer Einheit (Benutzung der originalen oder einer manipulierten Version der Systemsoftware) basiert [OKR05]. Die Grundideen zum normativen Systementwurf und der Bildung von Typglauben werden in Abschnitt 3.1.1 vorgestellt.

Gemäß Abschnitt 2.3.2 besteht die Schlüsselidee von Ansätzen mit Dritten zur Konfliktlösung darin, das Verhalten der Einheiten durch den Einsatz von nicht-abstreitbaren Marken für den vertrauenswürdigen Dritten *nachvollziehbar* zu machen. Diese *Beweismittel* unterstützen den Dritten darin, Betrugsverhalten zu erkennen und den davon verursachten Schaden rückgängig zu machen. Der eigene Ansatz baut auf dieser Idee auf [Obr04b]. Um Beweismittel einsetzen zu können, sind jedoch einige Hürden zu nehmen. Einerseits können wegen der Abwesenheit eines vertrauenswürdigen Dritten die Einheiten ihr Verhalten nur untereinander nachvollziehbar machen. Auch die Schlussfolgerungen, die die Einheiten aus den verfügbaren Beweismitteln ziehen, können nicht von einem Dritten umgesetzt werden und müssen daher in die selbstorganisierende soziale Kontrolle einfließen. Diese Umfunktionierung von Beweismitteln führt im eigenen Ansatz zum Entwurf von zweckbestimmten Beweismitteln, auf die das Empfehlungsmodell der verteilten Vertrauensbildung basiert. Die Grundideen zur Nachvollziehbarkeit durch Beweismittel werden in Abschnitt 3.1.2 vorgestellt.

[1] Der Begriff des normativen Systementwurfs ist am Konzept der Normen aus der Soziologie angelehnt [Tuo95]. Seine Ausrichtung unterscheidet sich von derjenigen der normativen Spieltheorie [FT91]. Diese ist darum bemüht vorauszusagen, wie sich ein rationaler Spieler in einer bestimmten Situation verhalten wird.

Teil II dieser Arbeit befasst sich mit den Einzelheiten des Entwurfs und führt dazu die Grundideen weiter aus.

3.1.1 Normativer Systementwurf und Bildung von Typglauben

Im Folgenden werden die Grundideen zum normativen Systementwurf und zur Bildung von Typglauben vorgestellt. Die Einzelheiten dazu werden im Kapitel 5 beziehungsweise im Kapitel 6 dieser Arbeit besprochen.

Normativer Systementwurf. Wir unterscheiden zwei *Typen* von Einheiten abhängig davon, welche Systemsoftware sie verwenden: Benutzt ein menschlicher Prinzipal die originale Systemsoftware, so ist seine Einheit *normativ*. Wenn hingegen eine manipulierte Version der Systemsoftware installiert ist, so nennen wir seine Einheit *strategisch*.

Der Entwurf des eigenen Ansatzes zur Betrugsvermeidung ist ein wesentlicher Teil des Entwurfs der originalen Systemsoftware. Es stellt sich also die Frage, welche Eigenschaften ein solcher Systementwurf haben muss, um unser Ziel eines existenzfähigen Informationssystems zu erreichen. Dazu wird eine Antwort in der folgenden Grundidee gegeben:

Grundidee 1:
Der Entwurf der originalen Systemsoftware orientiert sich an zwei *Maximen*:

- Das vorgeschriebene Verhalten ist *hinreichend kooperativ*, dass ein Informationssystem mit überwiegend normativen Einheiten einen Mehrwert für menschliche Benutzer bietet.
- Das vorgeschriebene Verhalten ist *hinreichend vorteilhaft*, dass normative Einheiten keinen Anreiz haben, strategisch zu werden.

Die zweite Maxime fordert, dass der menschliche Prinzipal einer normativen Einheit keinen Vorteil darin sieht, eine manipulierte Version der Systemsoftware zu verwenden. Jedoch handelt eine solche manipulierte Version eher im Sinne eines menschlichen Prinzipals, da sie in der Lage ist, die Vorteilhaftigkeit von Betrugsverhalten auszunutzen. Die einzige Möglichkeit, die zweite Maxime zu verfolgen, scheint daher darin zu liegen, dass der Systementwurf Betrugsverhalten für normative Einheiten vorsieht. Damit wird jedoch die Forderung der ersten Maxime verletzt, dass normative Einheiten sich hinreichend kooperativ verhalten. Die zweite Maxime scheint daher in Konflikt mit der ersten Maxime zu stehen.

Umgekehrt lässt sich auch die erste Maxime nicht verfolgen, ohne dass die zweite Maxime verletzt wird: Die erste Maxime zielt darauf, dass die Teilnahme am Informationssystem für den menschlichen Benutzer von Nutzen ist. Das ist nur dann der Fall, wenn kooperatives Verhalten im System vorherrscht. Zwar lässt sich die erste Maxime dadurch einhalten, dass im Systementwurf altruistisches Verhalten vorgeschrieben wird. Dies steht aber im direkten Widerspruch mit der zweiten Maxime, da sich in der Anwesenheit von Altruisten das Betrugsverhalten strategischer Einheiten besonders lohnt. Wir erhalten also, dass beide Maximen im Konflikt zu stehen scheinen.

Wie kann der Systementwurf dennoch beide Maximen gleichzeitig berücksichtigen? Es zeigt sich, dass das Kriterium der zweiten Maxime genauer untersucht werden muss. Dazu führen wir zwei Begriffe ein:

- *Manipulationskosten:* Durch das Erstellen, die Verbreitung und die Benutzung einer manipulierten Version der Systemsoftware entstehen für den menschlichen Prinzipal der Einheit

Kosten. Diese werden von *technischen* und *rechtlichen* Hindernissen verursacht. **(1)** Die Erstellung einer manipulierten Version oder die Beschaffung einer solchen von jemandem, der eine manipulierte Version bereitstellt, ist zeitaufwändig. **(2)** Es ist unklar, ob die Benutzung der manipulierten Version Vorteile für den menschlichen Prinzipal bringt. Dies gilt besonders dann, wenn die verwendete Version nicht selbst erstellt, sondern von anderen Benutzern übernommen wird. **(3)** Die Erstellung, Verbreitung und Benutzung einer manipulierten Version verstoßen gegen das Urheber- und Vertragsrecht.

- *Normativitätskosten:* Wenn im Systementwurf kooperatives Verhalten vorgesehen ist, entstehen für die normativen Einheiten Opportunitätskosten, da sie auf die Vorteilhaftigkeit von Betrugsverhalten verzichten müssen. Diese Kosten nennen wir Normativitätskosten.

Mit Hilfe dieser beiden Kostenarten lässt sich die zweite Maxime verfeinern. Dies führt uns zu folgender Grundidee:

Grundidee 2:
Ein *Kompromiss* zwischen den Maximen des Systementwurfs *wird ermöglicht* durch die folgenden Punkte:

- Eine normative Einheit wird *nur dann* strategisch, wenn die *Manipulationskosten geringer* sind als die *Normativitätskosten*.
- Die *Manipulationskosten* sind *nicht unerheblich*.
- Die *Normativitätskosten* werden dadurch *minimiert*, dass die Normen unter Zuhilfenahme der selbstorganisierenden sozialen Kontrolle so *selbstdurchsetzend* wie möglich sind.

Die Punkte der Grundidee werden im Folgenden genauer erklärt: *Der erste Punkt* stellt ein Kriterium dafür auf, dass normatives Verhalten hinreichend vorteilhaft ist, wie in der zweiten Maxime gefordert ist. Der menschliche Prinzipal einer normativen Einheit entscheidet sich nur dann zur Verwendung einer manipulierten Version der Systemsoftware, wenn dies für ihn einen Vorteil bringt. Dies ist genau dann der Fall, wenn die Normativitätskosten die Manipulationskosten übersteigen. Um dies zu verhindern, müssen die Manipulationskosten so hoch und die Normativitätskosten so niedrig wie möglich sein. *Der zweite Punkt* hält fest, dass aus technischen und rechtlichen Gründen die Manipulationskosten nicht zu vernachlässigen sind.

Es bleibt also noch die Frage, wie die Höhe der Normativitätskosten unter die der Manipulationskosten gedrückt werden kann. Dazu führt *der dritte Punkt* die selbstorganisierende soziale Kontrolle der verteilten Vertrauensbildung an. Die Normen des Systementwurfs sind genau dann selbstdurchsetzend, wenn für strategische Einheiten deren Befolgung im eigenen Interesse ist. Das Verhalten der normativen Einheiten muss also auf das der strategischen Einheiten abstrahlen. Dies wird durch zwei *Entwurfsprinzipien* erreicht:

1. Strategische Einheiten wollen *als normativ wahrgenommen* werden, um in vielen Transaktionen teilnehmen zu können. Zu diesem Zweck schreibt der Systementwurf für normative Einheiten vor, dass sie bevorzugt mit als normativ wahrgenommenen Einheiten in die Transaktionen gehen. Hierfür ist also eine *Glaubensbildung* über den *Typ* anderer Einheiten nötig.

2. Die einzige Möglichkeit, als normativ wahrgenommen zu werden, liegt darin, sich *normativ zu verhalten*. Dazu müssen die Einheiten in der Lage sein, ihre Erfahrungen mit bisherigen Transaktionspartnern *glaubhaft* an andere Einheiten zu kommunizieren. Dazu ist also ein Empfehlungssystem nötig, das Erfahrungen *nachvollziehbar* macht.

Diese Prinzipien legen die Ausrichtung des Entwurfs der verteilten Vertrauensbildung fest. Die folgenden Grundideen befassen sich daher mit der Umsetzung der von den Prinzipien geforderten Punkte, nämlich der Bildung von Typglauben und der Nachvollziehbarkeit in Empfehlungen.

Bildung von Typglauben. Wie wir in Abschnitt 2.4 gesehen haben, sehen existierende Ansätze der verteilten Vertrauensbildung eine rein verhaltensorientierte Glaubensbildung vor. Dadurch können erhaltene Empfehlungen und der Kontext des Transaktionsverhaltens nicht sinnvoll im Glaubensmodell berücksichtigt werden. Hinzu kommt die Forderung des ersten Entwurfsprinzips, dass Glauben über den Typ anderer Einheiten gebildet werden muss. Kann es eine anders geartete (nicht nur am Verhalten orientierte) Glaubensbildung geben, die diese Probleme beseitigt? Die folgende Grundidee ermöglicht es uns, diese Frage zu bejahen:

Grundidee 3:
Glaubensbildung *orientiert* sich am *Typ* anderer Einheiten. Um das *Verhalten* einer Einheit mit ihrem *Typ* in Beziehung zu setzen, werden folgende Möglichkeiten berücksichtigt:

- *Strategische Einhaltung:* Auch strategische Einheiten können die Absicht haben, Normen zu befolgen. Dies tun sie *kontextabhängig*.
- *Unbeabsichtigtes Betrugsverhalten:* Sowohl normative als auch strategische Einheiten können sich trotz kooperativer Absichten fehlverhalten.

Das Glaubensmodell muss Verhalten und Typ einer Einheit aus zweierlei Gründen in Beziehung setzen. Einerseits ist dies die Voraussetzung dafür, dass Verhaltensinformation für die Revision des Typglaubens probabilistisch richtig berücksichtigt werden kann. Andererseits muss aus dem Typglauben eine Einschätzung wahrscheinlichen zukünftigen Verhaltens ableitbar sein, damit Vertrauensentscheidungen getroffen werden können.

Die Beziehung wird in zwei Schritten hergestellt: **(1)** Wie hängt der Typ einer Einheit mit ihrer *Absicht* zu normativem Verhalten zusammen? Auch strategische Einheiten können diese Absicht haben, vor allem wenn sie erst Vertrauen aufbauen wollen, um anschließend effektiver betrügen zu können. Diese Möglichkeit strategischer Einhaltung wird im Modell der Glaubensbildung explizit mit aufgenommen. Gemäß dem Modell sind die Absichten strategischer Einheiten abhängig vom Transaktionskontext. Bei hohem Transaktionswert wird bevorzugt betrogen, da dies besonders lohnenswert ist. **(2)** Wie hängt die Absicht einer Einheit mit ihrem Verhalten zusammen? Wie in Abschnitt 1.2.2 bereits festgestellt, können selbst kooperative Absichten zu Verhalten führen, das der Transaktionspartner als Betrugsverhalten wahrnimmt. Daher berücksichtigt das Glaubensmodell die Möglichkeit unbeabsichtigten Betrugsverhaltens.

Die sich ergebende Glaubensbildung realisiert nicht nur das erste Entwurfsprinzip. Darüber hinaus wird die Grundlage für eine probabilistisch korrekte Berücksichtigung von Empfehlungsinformation gelegt. Außerdem ermöglicht die Typorientierung der Glaubensbildung den Einsatz von sozialen Beweismitteln, wie im nächsten Abschnitt beschrieben.

3.1.2 Einsatz von nicht-abstreitbaren Beweismitteln

Im Folgenden werden die Grundideen zum Einsatz von nicht-abstreitbaren Marken *(Beweismitteln)* vorgestellt. Wir unterscheiden zwischen transaktionalen und sozialen Beweismitteln. Die Einzelheiten dazu werden im Kapitel 7 beziehungsweise im Kapitel 8 dieser Arbeit besprochen.

Konzept der Beweismittel. Ein *Beweismittel* ist eine nicht-abstreitbare Marke, durch deren Ausstellung sich eine Einheit an eine *Aussage* über eine andere Einheit bindet. Jede andere Einheit ist in der Lage, bei Erhalt des Beweismittels zu erkennen, wer das Beweismittel ausgestellt hat und zu welcher Aussage er sich festgelegt hat. In dieser Hinsicht kann einer beliebigen Einheit diese Festlegung *glaubhaft* gemacht werden. Dazu ist lediglich das Vorzeigen des entsprechenden Beweismittels notwendig. Gerade in Informationssystemen mit beschränkten Kommunikationsmöglichkeiten wie im Ad-hoc Netz des Campus-Szenarios ist diese Eigenschaft von Vorteil: Die Aussage einer Einheit kann nämlich auch in ihrer *Abwesenheit* nach Bedarf glaubhaft reproduziert werden.

Das Kernproblem existierender Ansätze zur verteilten Vertrauensbildung liegt darin, dass ein Empfehler freie Wahl über den Inhalt der auszustellenden Empfehlung hat. Damit wird das Empfehlen selber zum *billigen Gerede* (engl.: cheap talk [BG97]), dessen Inhalt keinen Anhaltspunkt über tatsächlich gezeigtes Verhalten bietet. Um dieses Problem zu lösen, bedarf es einer Möglichkeit, in Empfehlungen eigene Beobachtungen glaubhaft *signalisieren* zu können. Folgende Grundidee schlägt hierfür den Einsatz von Beweismitteln vor:

> **Grundidee 4:**
> Das Verhalten einer Einheit wird durch den Einsatz von Beweismitteln für andere Einheiten *nachvollziehbar* gemacht. *Empfehlungen* müssen also durch entsprechende Beweismittel *gestützt* werden.

Aus dieser Grundidee ergeben sich hauptsächlich drei Konsequenzen für die Konzeption der verteilten Vertrauensbildung:

- Es sind Arten von Beweismitteln zu finden und einzusetzen, deren Aussagen zu Nachvollziehbarkeit bezüglich des Verhaltens einer Einheit führen.
- Das *Empfehlungsmodell* muss sich auf diese Beweismittel stützen. Dadurch muss ein vollkommen neues Empfehlungssystem entworfen werden.
- Beweismittel und Empfehlungen zielen letztendlich darauf, dass die Erfahrungen einer Einheit in die Glaubensbildung anderer Einheiten mit einfließen. Dazu ist festzulegen, wie der *Typglauben* einer Einheit bei Erhalt einer Empfehlung zu *revidieren* ist.

Im verbleibenden Teil des Entwurfs ziehen wir diese drei Konsequenzen entlang zweierlei Arten von Beweismitteln, nämlich *transaktionalen* und *sozialen* Beweismitteln.

Transaktionale Beweismittel. Im Laufe einer Transaktion kann eine Einheit ihren Transaktionspartner betrügen, indem sie ihm die versprochene Aktion vorenthält. Um Betrugsverhalten nachvollziehbar zu machen, bedarf es also Beweismittel, die im Zuge der Transaktion ausgetauscht werden. Der Aussteller von solchen *transaktionalen Beweismitteln* bindet sich an eine Aussage über das Verhalten seines Transaktionspartners. Wir unterscheiden drei Arten von transaktionalen Beweismitteln:

- *Vertrag:* Vor dem Ausführen der Aktionen stellen sich die Transaktionspartner jeweils gegenseitig einen Vertrag aus, in dem sie versprechen, welche Aktion sie im Zuge der Transaktion ausführen.
- *Quittung:* Nach dem Ausführen der Aktionen werden untereinander Quittungen ausgestellt. Der Aussteller einer Quittung bestätigt seinem Transaktionspartner, dass er seine Versprechungen (wie sie im Vertrag festgehalten sind) eingehalten hat.

- *Negative Empfehlung:* Kommt es im Laufe der Transaktion zu Betrugsverhalten, so kann eine Einheit ihren Transaktionspartner anderen Einheiten gegenüber negativ empfehlen. Eine solche negative Empfehlung bezieht sich dabei jeweils auf eine Transaktion.

Wie wirkt sich der Einsatz dieser drei Arten transaktionaler Beweismittel auf das Empfehlungsmodell und die Glaubensbildung aus? Dazu müssen wir uns zunächst fragen, unter welchen Umständen aus der Sicht des Gesamtsystems das Ausstellen negativer Empfehlungen wünschenswert ist. Kommt es im Laufe einer Transaktion zwischen zwei Einheiten zu Betrug, so lässt sich für außenstehende Einheiten nicht erkennen, von wem dieser Betrug ausging. Wir reden daher von einem *Konflikt* zwischen den Transaktionspartnern, wenn einer der beiden im Laufe der Transaktion betrogen hat. Unser Ziel muss es also sein, dass alle Konflikte den außenstehenden Einheiten durch das Ausstellen von negativen Empfehlungen mitgeteilt werden. Dadurch wird eine Einheit, die wiederholt betrügt, identifizierbar. Die folgende Grundidee zeigt, wie ein solches Empfehlungsverhalten selbstdurchsetzend gemacht wird:

Grundidee 5:
Eine Einheit wird *genau dann* negativ empfohlen, wenn sie an einer Transaktion teilnahm, in der es zu Betrug kam. Dies wird durch folgende Punkte ermöglicht:

- Jede Einheit kann *nur* einen früheren Transaktionspartner negativ empfehlen, da dies dessen *Vertrag* erfordert.
- Negative Empfehlungen über Transaktionen *ohne Betrug* werden *nicht* ausgestellt, da sie durch die *Quittung* widerlegbar sind.
- Jede Einheit empfiehlt negativ *wann immer möglich*, um den *komparativen Vorteil*, als Erster zu empfehlen, zu nutzen.

Der *logische Zusammenhang* zwischen den drei Punkten ist folgendermaßen: Der erste Punkt stellt sicher, dass negative Empfehlungen sich immer auf Transaktionen beziehen, die auch tatsächlich stattgefunden haben. Damit ist es (im Gegensatz zu den existierenden Ansätzen der verteilten Vertrauensbildung) nicht möglich, eine Einheit zu diffamieren, ohne mit ihr in eine Transaktion gegangen zu sein. Der zweite Punkt schränkt die Menge der möglichen negativen Empfehlungen weiter ein. Eine Einheit kann also nur dann negativ empfohlen werden, wenn sie in einer Transaktion teilgenommen hat, in der es zu Betrug kam. Der dritte Punkt stellt sicher, dass in einem solchen Fall auch tatsächlich negativ empfohlen wird. Damit erhalten wir das erwünschte Ergebnis, dass eine Einheit genau dann negativ empfohlen wird, wenn sie an einem Konflikt beteiligt ist.

Der zweite Punkt der Grundidee besagt, dass eine Einheit ihren Transaktionspartner nicht negativ empfiehlt, nachdem sie ihm eine Quittung ausgestellt hat. Dies wird durch eine Erweiterung des Empfehlungsmodells sichergestellt: Jede Einheit ist in der Lage, in einer *Selbstempfehlung* die Quittungen, die sie von ihren früheren Transaktionspartnern erhalten hat, an andere Einheiten weiterzugeben. Wenn eine Einheit nach dem Ausstellen einer Quittung dennoch ihren Transaktionspartner negativ empfiehlt, ist also ihr *inkonsistentes* Verhalten erkennbar. Die Folgekosten sind für diese Einheit deswegen besonders hoch, weil ihr inkonsistentes Verhalten von jeder Einheit durch die Nichtabstreitbarkeit der Quittung und der negativen Empfehlung nachvollzogen werden kann. Wir fassen also zusammen, dass es nach dem Ausstellen einer Quittung deswegen nicht zu einer negativen Empfehlung durch den Transaktionspartner kommt, weil er dadurch von allen anderen Einheiten als *strategische Einheit* wahrgenommen werden würde.

Der *dritte Punkt* der Grundidee verlangt, dass negative Empfehlungen derart in der Glaubensbildung berücksichtigt werden, dass es einen Anreiz für ihre Ausstellung gibt. Daher wird folgender Ansatz für die Glaubensrevision verfolgt: Nach Erhalt einer negativen Empfehlung wird der Empfohlene entsprechend der Glaubwürdigkeit des Empfehlers abgewertet. Der Typglauben über den Empfehler selbst bleibt jedoch unverändert. Empfehlen sich zwei Einheiten gegenseitig negativ, so wird also diejenige Einheit weniger abgewertet, die als erste empfohlen hat. Dies liegt daran, dass bei Erhalt der Empfehlung der zweiten Einheit diese bereits abgewertet worden ist und damit ihre Glaubwürdigkeit eingebüßt hat. Eine Untersuchung des *Spiels negativer Empfehlungen* zeigt, dass dieser *komparativer Vorteil*, als erster zu empfehlen, dafür ausreicht, dass jede Einheit wann immer möglich negativ empfiehlt.

In der Glaubensbildung werden Selbstempfehlungen anders als negative Empfehlungen behandelt. Das Vorzeigen von Quittungen in einer Selbstempfehlung dient alleine zur Widerlegung von inkonsistenten negativen Empfehlungen. Eine weitergehende (nämlich aufwertende) Glaubensrevision basierend auf Quittungen ist nicht sinnvoll, da sie auf einfache Weise durch *verschwörerisches* (engl.: collusive) Verhalten ausgenutzt werden könnte. Es bedarf also einer anderen Art von Beweismittel, mit deren Hilfe eine Einheit in der Lage ist, sich positiv zu empfehlen. Der Rest dieses Abschnitts befasst sich mit solchen Beweismitteln.

Soziale Beweismittel. Es zeigt sich, dass außer den transaktionalen Beweismitteln noch eine Art von Beweismitteln benötigt wird, mit der sich auch der Glauben einer Einheit über den Typ anderer Einheiten festhalten lässt. Solche *sozialen Beweismittel* sind losgelöst von einzelnen Transaktionserfahrungen und werden abhängig vom Typglauben ausgestellt.

Eine Analyse der Möglichkeiten sozialer Beweismittel ergibt, dass diese Beweismittel die Form von *Bürgschaften* haben sollten. Genauer gesagt legt sich der Aussteller des Beweismittels (der *Bürge*) fest, dass er eine andere Einheit (der *Gebürgte*) als normativ ansieht. Damit macht der Bürge seinen Typglauben anderen Einheiten gegenüber explizit. Da der Bürge keine Möglichkeit zum Widerruf hat, ist die Gültigkeit einer Bürgschaft zeitlich begrenzt.

Welcher Vorteil entsteht aus dem Einsatz solcher Bürgschaften? Folgende Grundidee gibt hierauf eine Antwort.

Grundidee 6:
Normative Einheiten können sich durch *Selbstempfehlungen* glaubhaft von strategischen Einheiten *abgrenzen*. Dies wird durch folgende Punkte ermöglicht:

- Jede Einheit kann sich selbst empfehlen, indem sie ihre eigenen *Bürgen* auflistet.
- Eine Bürgschaft wird zwischen einem Paar von Einheiten nur unter *beidseitigem Einverständnis* abgeschlossen.
- Jede Einheit ist *nur* auf Bürgschaften mit *normativen Einheiten* aus, da die Einschätzung der Normativität einer Einheit
  - vom *Typglauben* gegenüber ihrer Bürgen *abhängt* und
  - durch das *Verhalten* ihrer Bürgen *beeinflusst wird*.

Auch bei dieser Grundidee verdeutlichen wir den *logischen Zusammenhang* zwischen ihren Punkten: Gemäß dem zweiten Punkt schließen Einheiten immer *beidseitig* Bürgschaften ab. Damit ist der Bürge auch gleichzeitig der Gebürgte und umgekehrt. Nun fügen wir die Aussage

vom dritten Punkt hinzu, dass jede Einheit nur mit normativen Einheiten eine solche beidseitige Bürgschaftsbeziehung eingehen möchte. Als Folgerung erhalten wir, dass Bürgschaften nur unter normativen Einheiten zustande kommen. Gemäß dem ersten Punkt können in Selbstempfehlungen eigene Bürgen (und ihre Bürgschaften) aufgelistet werden. Der Aussteller einer solchen Selbstempfehlung ist also in der Lage, durch die Zahl und der Qualität eigener Bürgen seine eigene Normativität glaubwürdig zu kommunizieren. Wir erhalten also, dass sich normative Einheiten glaubhaft von strategischen Einheiten abgrenzen können.

Im Folgenden gehen wir genauer auf den *dritten Punkt* ein, da er den Kern der Grundidee darstellt. Es gibt zwei Gründe, warum eine Einheit mit einer anderen Einheit eine Bürgschaftsbeziehung eingehen will:

- *Bürgschaft als Signal:* Kommt die Bürgschaftsbeziehung zustande, so erhalten beide Partner der Beziehung jeweils eine Bürgschaft voneinander. Mit dieser Bürgschaft sind Selbstempfehlungen möglich. Wie werden diese aber von außenstehenden Einheiten bewertet? Je normativer der Bürge erscheint, desto glaubwürdiger ist seine Festlegung, dass der Gebürgte normativ ist, und desto stärker wird damit der Gebürgte aufgewertet. Daraus folgern wir, dass jede Einheit solche Bürgen sucht, die normativ erscheinen.

- *Bürgschaft als Investition:* Bürgschaften dienen nicht nur als Signal bei Selbstempfehlungen. Darüber hinaus hat das *Verhalten* des Gebürgten auch Auswirkungen auf den Glauben über den Bürgen. Wenn sich die Festlegung des Bürgen, dass der Gebürgte normativ ist, in Transaktionen mit dem Gebürgten bewahrheitet (beziehungsweise sich als falsch herausstellt), wird auch der Bürge normativer (beziehungsweise strategischer) als zuvor wahrgenommen. In diesem Sinne stellt das Ausstellen einer Bürgschaft eine Investition dar, die sich umso mehr auszahlt je normativer sich der Gebürgte verhält. Auch hier erhalten wir, dass jede Einheit Bürgschaften mit solchen Einheiten sucht, die sich normativ verhalten.

Bei beiden Gründen ergibt sich also ein Vorzug von Bürgschaftsbeziehungen mit normativen Einheiten.

Wie wird die Ausrichtung von Bürgschaften als Signale und Investitionen in der Glaubensbildung umgesetzt? Wie bereits angedeutet muss jede Einheit Informationen über die Bürgschaften Anderer in ihre Glaubensbildung einbeziehen. Außer dem *individuellen* Typglauben, wie er bisher betrachtet wurde, existiert damit noch ein *sozialer* Typglaube. Dieser setzt sich aus dem individuellen Typglauben über die betrachtete Einheit und ihre Bürgen zusammen. Da Bürgschaften zeitlich begrenzt sind und damit ungültig werden können, ist nur der individuelle Typglaube explizit zu speichern und der soziale Typglaube nach Bedarf daraus abzuleiten. Es werden Vorschriften zur Glaubensbildung und -revision vorgestellt, die dazu in der Lage sind.

3.2 Grundideen der Evaluation

Der Entwurf des eigenen Ansatzes zielt letztendlich auf die Verifikation unserer These, dass die Vision von Informationssystemen wie die des Campus-Szenarios realisiert werden kann. Um diese These validieren zu können, ist eine quantitative Bewertung des entworfenen Ansatzes notwendig. Zu diesem Zweck wird der eigene Ansatz *simulativ evaluiert*.

Im Folgenden geben wir eine Übersicht der Grundideen, die bei der Evaluation zum Einsatz kommen. Teil III dieser Arbeit befasst sich mit den Einzelheiten der Evaluation. Dazu werden die Grundideen in Kapitel 9 weiter ausgeführt. Die Durchführung der Evaluation und ihre Ergebnisse sind in Kapitel 10 zu finden.

Ausrichtung der Evaluation. Die beiden Maximen des Entwurfs fordern, dass normatives Verhalten zugleich hinreichend kooperativ und hinreichend vorteilhaft ist. Alle Teile des eigenen Ansatzes sind an diesen Maximen orientiert. Was fehlt, ist eine quantitative Untersuchung, unter welchen *Rahmenbedingungen* diese beiden Kriterien erfüllt werden. Für die Rahmenbedingungen, in denen dies der Fall ist, kann unsere These validiert werden. Die folgende Grundidee zeigt, wie dieser Gedankengang Eingang in die Ausrichtung der Evaluation findet:

> **Grundidee 7:**
> Die These kann in solchen Rahmenbedingungen *validiert* werden, in denen die Evaluation des Ansatzes zeigt, dass menschliche Prinzipale von normativen Einheiten *keinen* Anreiz haben,
>
> - aus dem System *auszutreten* oder
> - ihre Einheit zu *manipulieren*.

Die Verbindung der beiden Punkte mit den beiden Maximen des Entwurfs ist wie folgt: Wenn normative Einheiten sich hinreichend kooperativ verhalten, bietet die Teilnahme am Informationssystem einen Mehrwert derart, dass ihre menschlichen Prinzipale nicht aus ihm austreten wollen. Zur Bewertung dieses Kriteriums ist der *Individualnutzen* normativer Einheiten zu erfassen. Wenn normatives Verhalten hinreichend vorteilhaft ist, führt die Benutzung einer manipulierten Version der Systemsoftware zu einem zu geringen Vorteil. Eine Beurteilung davon ist möglich, indem die *Normativitätskosten* gemessen und mit den *Manipulationskosten* verglichen werden. Für den Fall, dass beide Kriterien erfüllt sind, ist das Informationssystem existenzfähig und unsere These kann damit für die untersuchten Rahmenbedingungen validiert werden.

Um zu den erforderlichen Messergebnissen zu kommen, ist eine Werkzeugunterstützung unabdingbar. Als Basis der Evaluation dient das Simulationsrahmenwerk *DIANEmu*, das eine wirklichkeitsgetreue Simulation des Campus-Szenarios ermöglicht. Darauf aufbauend werden zwei Werkzeuge entwickelt:

- *Integration in simulatives Rahmenwerk:* Der eigene Ansatz wird in das Rahmenwerk der Simulationsumgebung von DIANEmu integriert. Als Ergebnis entsteht ein Werkzeug, das für vorgegebene Rahmenbedingungen die erforderlichen Messergebnisse liefert. Die dabei entstehende Simulationsumgebung nennen wir simulatives Kooperationsturnier.

- *Erweiterung um Sensibilitätsanalyse:* Für die Verifikation der These muss der eigene Ansatz unter einer Vielzahl von verschiedenen Rahmenbedingungen evaluiert werden[2]. Zu diesem Zweck wird ein Werkzeug entwickelt, das die *Sensibilität* der Messergebnisse in Abhängigkeit unterschiedlicher Rahmenbedingungen automatisiert analysiert.

Berücksichtigung der Möglichkeiten zur Manipulation. Bei der simulativen Evaluation des eigenen Ansatzes ergibt sich die Frage, wie das Verhalten strategischer Einheiten zu modellieren ist. Diese Frage lässt sich darauf zurückführen, welche manipulierten Versionen der Systemsoftware in einem realen Informationssystem erstellt werden und für welche dieser Versionen sich ein menschlicher Prinzipal, der zur Manipulation bereit ist, entscheidet. Eine Beantwortung dieser Frage ist keineswegs trivial, da hierzu die Angriffspunkte gegen das normative Verhalten,

[2] Dies wird von allen existierenden Ansätzen der verteilten Vertrauensbildung vernachlässigt. Sie werden nur für eine geringe Zahl als besonders relevant erachteter Rahmenbedingungen evaluiert.

wie es im eigenen Ansatz vorgesehen ist, systematisch aufgedeckt werden müssen[3]. Die folgende Grundidee zeigt, wie hierfür vorgegangen werden muss:

Grundidee 8:
Mögliche Richtungen der *Manipulation* werden zwecks realitätsnaher Simulation *antizipiert*. Dies wird durch die folgenden Punkte ermöglicht:

- Erfolg versprechende *Gegenstrategien* werden von *Versuchspersonen* in einer *interaktiven* Simulationsumgebung *aufgezeigt*.
- Ein simulativer *Vergleich* zeigt, *welche* Gegenstrategien für bestimmte Rahmenbedingungen am besten abschneiden und daher *zu erwarten* sind.

In einem *ersten Schritt* werden also Erfolg versprechende Richtungen der Manipulation *aufgefunden*. Dieses Auffinden ist insofern *systematisch*, als eine größere Zahl von Versuchspersonen beteiligt ist, die sich mit dem Gesamtsystem auskennen und denen beim Finden von Gegenstrategien keinerlei Einschränkungen auferlegt sind. Um dies zu ermöglichen, wird ein Werkzeug, das *interaktive Kooperationsturnier*, entwickelt, mit dem jede Versuchsperson aus der Sicht einer Einheit am Gesamtsystem teilnehmen und ihr Verhalten bestimmen kann. Dadurch ergeben sich folgende Vorteile:

- Für die Versuchspersonen sind keinerlei technische Vorkenntnisse (wie etwa Programmierkenntnisse) erforderlich. Die Vorgänge im System werden graphisch ansprechend aufgearbeitet, so dass jede Versuchsperson von der jeweiligen Situation Bescheid weiß und das Verhalten ihrer eigenen Einheit in ihrem Sinne lenken und bestimmen kann. Dadurch erweitert sich der Kreis derer, die sich am Auffinden der Gegenstrategien beteiligen können.
- Die Versuchspersonen entwickeln sozusagen spielerisch ihre Strategien und müssen sich nicht *a priori* auf sie festlegen. Nach jedem Lauf des interaktiven Kooperationsturniers werden die erfolgreichsten Versuchspersonen nach ihren Strategien befragt. Somit werden die Gegenstrategien *a posteriori* definiert. Diese Vorgehensweise führt auch dazu, dass die Versuchspersonen zur Teilnahme am interaktiven Kooperationsturnier und zum Finden erfolgreicher Gegenstrategien *motiviert* sind.

In einem *zweiten Schritt* sind die aufgefundenen Gegenstrategien zu *bewerten*. Diese Bewertung zielt darauf, eine Aussage zu treffen, unter welchen Rahmenbedingungen welche Gegenstrategie am erfolgreichsten ist. Hierzu wird das simulative Kooperationsturnier bei unterschiedlichen Rahmenbedingungen durchlaufen. Damit sind wir in der Lage, bei der Gesamtevaluation die jeweils beste Gegenstrategie in Abhängigkeit der zu simulierenden Rahmenbedingungen für die strategischen Einheiten auszuwählen.

Die Realitätsnähe der Simulation wird weiter erhöht, indem strategische Einheiten bei der Wahl ihrer Gegenstrategie die Rahmenbedingungen auch leicht verzerrt wahrnehmen können. Dies lässt sich an einem Beispiel verdeutlichen: Bei der Wahl der Gegenstrategie ist die Zahl der zu erwartenden Transaktionsgelegenheiten ein wichtiger Faktor. Es wäre aber unrealistisch anzunehmen, dass die Einheiten diese Zahl genau kennen. Dadurch, dass manche Einheiten diese Zahl unter- beziehungsweise überschätzen, unterscheiden sich die strategischen Einheiten bei ihrer Wahl der Gegenstrategie. Wir erhalten also eine *Heterogenisierung* der verfolgten Gegenstrategien dadurch, dass die Rahmenbedingungen *unpräzise wahrgenommen* werden.

[3]Der Stand der Technik hierzu ist, dass der Systementwerfer selbst Angriffspunkte durch eigenes Nachdenken identifiziert. Diese Vorgehensweise ist weder systematisch noch führt sie zu glaubwürdigen Evaluationsergebnissen.

3.3 Zusammenfassung

In diesem Kapitel wurde vorgestellt, wie diese Arbeit vorgeht, um unserer These über die Existenzfähigkeit von Informationssystemen wie im Campus-Szenario validieren zu können. Dazu wurden die Grundideen für den Entwurf und die Evaluation eines eigenen Ansatzes zur verteilten Vertrauensbildung besprochen.

Der *Entwurf* des Ansatzes orientiert sich an zwei Maximen, deren Zielkonflikt durch die Existenz technischer und rechtlicher Hindernisse gegen die Manipulation gelöst wird. Durch eine probabilistisch fundierte Glaubensbildung, die sich am Typ der Einheiten orientiert, wird die Grundlage zur verteilten Vertrauensbildung gelegt. Zur glaubwürdigen Verteilung von Transaktionserfahrungen werden Beweismitteln benutzt. Transaktionale Beweismittel erlauben die Identifikation strategischer Einheiten, während sich Einheiten mit Hilfe sozialer Beweismittel als normativ ausweisen können.

In der simulativen *Evaluation* wird der Ansatz quantitativ bewertet, um die These validieren zu können. Aus den beiden Maximen des Entwurfs werden dazu zwei Evaluationskriterien abgeleitet. Als Simulationswerkzeug steht DIANEmu zur Verfügung. Die Realitätsnähe der Simulation wird weiter gesteigert, indem die Möglichkeiten zur Manipulation aus einer interaktiven Simulation gewonnen werden und somit antizipiert werden können.

In den folgenden Teilen II und III wird der Ansatz dieser Arbeit im Detail vorgestellt.

Teil II

Entwurf

Kapitel 4

勢者 因利而制權也

"Derjenige, der mit jedweder Situation umgehen kann, zeichnet sich dadurch aus, dass er bei Eintreten eines unvorhergesehenen Ereignisses diejenige Maßnahme ergreift, die ihm den größten Vorteil verschafft."
(Sun-tze's Kunst der Kriegsführung, Strategie, 1.17)

Grundlagen

Teil II dieser Arbeit befasst sich mit dem Entwurf des eigenen Ansatzes. Hierfür sind zunächst die Ergebnisse der Forschungsbereiche vorzustellen, die diesem Entwurf zugrunde liegen. In diesem Kapitel werden diese Grundlagen eingeführt. Es handelt sich bei ihnen vor allem um die Konzepte der Spieltheorie, die in Abschnitt 4.1 behandelt werden. Die weiteren Grundlagen werden in Abschnitt 4.2 besprochen.

4.1 Spieltheorie

Die Spieltheorie befasst sich mit der Frage, wie sich autonome Einheiten, die ihren eigenen Vorteil suchen, unter vorgegebenen Rahmenbedingungen verhalten. Die Methoden und Aussagen der Spieltheorie haben also einen direkten Bezug zum Problemkreis dieser Arbeit.

Dieser Abschnitt befasst sich daher mit den Konzepten und Aussagen der Spieltheorie insoweit, als sie dieser Arbeit als Grundlage dienen. Abschnitt 4.1.1 führt in die Spieltheorie ein, indem ihre Grundzüge besprochen werden. Nachfolgend wird in Abschnitt 4.1.2 der Zweig der Spieltheorie vorgestellt, der sich mit Informationsasymmetrie und Signalisierung beschäftigt. Abschließend gehen wir in Abschnitt 4.1.3 auf die evolutionäre Spieltheorie ein.

4.1.1 Grundzüge

Im Folgenden werden die Grundzüge der Spieltheorie vorgestellt. Dazu gehen wir zunächst darauf ein, wie die Rahmenbedingungen autonomen Handelns als Spiel modelliert werden. Anschließend werden die Methoden vorgestellt, mit deren Hilfe ein solches Spiel analysiert werden kann. Eine umfassende Behandlung der Grundzüge der Spieltheorie findet sich in [FT91].

Das Spiel als Modell. In einem Spiel werden nur diejenigen Aspekte der Realität modelliert, die bei der Vorhersage des Verhaltens autonomer Einheiten von Relevanz sind. Für die einzelnen Bestandteile des Modells bedeutet dies Folgendes:

- *Handlungen:* Eine autonome Einheit ist zu unterschiedlichen Handlungen in der Lage. Manche Handlungen wirken sich lediglich auf die Einheit selbst aus. Solche Handlungen mit rein lokaler Auswirkung beeinflussen nicht andere Einheiten und sind daher für die spieltheoretische Analyse nicht von Belang. Aussagen darüber, wie Entscheidungen bezüglich solcher Handlungen zu treffen sind, werden im Rahmen der Entscheidungstheorie gemacht [Ber85]. Im Modell des Spiels sind also nur diejenigen Handlungen zu berücksichtigen, durch deren Ausführung andere Einheiten beeinflusst werden.

- *Spieler:* Eine Gruppe von Einheiten wird nur dann als Spieler in das Modell aufgenommen, wenn die Einheiten zu Handlungen fähig sind, die sich auf sie gegenseitig auswirken. Eine Einheit wird also nicht als Spieler modelliert, wenn ihre Handlungen auf keine andere Einheit Auswirkungen haben oder sie selbst von keiner der Handlungen Anderer einen Einfluss erfährt.

- *Ausgang:* Der Ausgang eines Spiels ist dadurch definiert, welche Handlungen die Spieler ausführen. Dabei ist zu unterscheiden, ob die Spieler *simultan* oder *sequentiell* handeln. Im letzteren Fall kann ein Spieler bei seiner Entscheidung über sein Handeln die vorigen Handlungen anderer Spieler berücksichtigen. Dadurch, dass in einem solchen sequentiellen Spiel die Spieler nacheinander handeln, sprechen wir auch vom *Ziehen* der Spieler.

- *Auszahlung:* Jeder Spieler hat für jeden möglichen Ausgang ein bestimmtes Nutzenniveau, das Auszahlung (engl.: payoff) genannt wird. In unserem Systemmodell von Informationssystemen wird der Nutzen einer Einheit durch den ihres menschlichen Prinzipals definiert. Um das Nutzenniveau einer Einheit zu bestimmen, ist zu berücksichtigen, welche Vor- und Nachteile der jeweilige Ausgang für ihren Prinzipal hat. Im Campus-Szenario ergeben sich die Vorteile vor allem aus dem Erhalt erwünschter Information, während die Nachteile aus der Bereitstellung eigener Ressourcen herrühren.

- *Strategie:* Einen Handlungsplan eines Spielers, der angibt, in welcher Spielsituation welche Handlung gewählt wird, nennen wir Strategie. In einem simultanen Spiel wird die Strategie eines Spielers also dadurch definiert, zu welcher Handlung er sich entscheidet. In sequentiellen Spielen ist die Festlegung einer Strategie umfangreicher, da hierbei für alle Kombinationen früherer Züge definiert sein muss, welche Handlung zu wählen ist.

Ein Spieler nimmt also an einem Spiel teil, indem er sich eine Strategie zurechtlegt. Die Spieler können aufgrund ihrer Autonomie nicht zum Wählen bestimmter Strategien gezwungen werden. Sie alleine treffen die Entscheidung über ihre Strategie. In der Literatur wird die Theorie zu solchen Spielen *unkooperative Spieltheorie* (engl.: noncooperative game theory) bezeichnet.

Die Beziehung zwischen den eingeführten Begriffen werden in folgendem Beispiel, dem *Gefangenendilemma* (engl.: prisoners' dilemma), verdeutlicht. Im Modell dieses Spiels wird festgehalten, wie zwei Spieler A und B untereinander in Konkurrenz stehen. Zur Illustration gehen wir davon aus, dass es sich bei den Spielern um benachbarte Ladenbesitzer handelt[1]. Ihr einziger Handlungsspielraum liegt darin, welche Preise sie für die Produkte festsetzen, die sie verkaufen. Dabei haben beide Spieler die Wahl zwischen hohen und niedrigen Preisen. Da sie diese Wahl simultan treffen müssen, ist ihre Strategie dadurch definiert, zu welcher Preissetzung sie sich unabhängig von ihrem jeweiligen Gegenspieler entscheiden. Wenn die Spieler dasselbe Preisniveau wählen, so

[1]Die erste Formalisierung der hier betrachteten Konkurrenzsituation findet sich in [Tuc50]. Zur Illustration verwendet sie die Situation zweier Gefangenen, die sich gegenseitig denunzieren können. Hieraus erklärt sich die Namensgebung des Dilemmas.

Tabelle 4.1: Das Gefangenendilemma

		Ladenbesitzer B	
		hoher Preis *(Kooperation)*	niedriger Preis *(Betrug)*
Ladenbesitzer A	hoher Preis *(Kooperation)*	$(3, 3)$	$(0, 5)$
	niedriger Preis *(Betrug)*	$(5, 0)$	$(1, 1)$

teilen sie sich den Markt gerecht auf. Bei einem beidseitig hohen Preisniveau führt dies zu einem höheren Gewinn als bei einem beidseitig niedrigen Preisniveau. Im Folgenden verdeutlichen wir dies mit einem Zahlenbeispiel: In den Auszahlungen der Ladenbesitzer schlagen sich gleich hohe Preise dadurch nieder, dass beide Spieler einen Nutzen von 3 bei hohen Preisen und von 1 bei niedrigen Preisen haben. Wählen die Ladenbesitzer allerdings ein unterschiedliches Preisniveau, unterscheiden sich ihre Auszahlungen. Der Ladenbesitzer mit niedrigen Preisen hat eine Auszahlung von 5, da er den Markt für sich alleine vereinnahmt. Der andere Ladenbesitzer verliert seine Kunden und erhält daher eine Auszahlung von 0.

In der Literatur zum Gefangenendilemma werden die Handlungsalternativen der in Konkurrenz stehenden Ladenbesitzer auch Kooperation (engl.: cooperation) und Betrug (engl.: defection) genannt. Die Motivation hierfür ist wie folgt: Entschließen sich beide Ladenbesitzer einvernehmlich zu hohen Preisen, so teilen sie den Markt gerecht auf und erhalten höhere Auszahlungen als bei beidseitig niedrigen Preisen. In dieser Hinsicht stellt das Festsetzen von hohen Preisen kooperatives Verhalten dem anderen Ladenbesitzer gegenüber dar. Weicht jedoch einer der Ladenbesitzer von dieser einvernehmlichen Kooperation durch das Festsetzen niedriger Preise ab, so bringt er den Anderen um seine Marktanteile. Diesbezüglich stellt sein Verhalten Betrug dar.

Die Beschreibung des Gefangenendilemmas wird in Tabelle 4.1 zusammengefasst. Dargestellt ist das Modell des Spiels in der *Normalform*. Auf der linken Seite stehen die Handlungsalternativen des Ladenbesitzers A, rechts die des Ladenbesitzers B. Für jeden Ausgang wird die Auszahlung der Spieler durch einen Vektor dargestellt, dessen erste Komponente die Auszahlung von A und dessen zweite Komponente die Auszahlung von B festhält. Entsprechend bezeichnen wir die Ausgänge mit der Kombination der Handlungen, die die Ladenbesitzer A und B jeweils wählen. Bei dem Ausgang *(niedriger Preis, hoher Preis)* erzielt Ladenbesitzer A zum Beispiel die höchste Auszahlung, während Ladenbesitzer B gleichzeitig die niedrigste Auszahlung erhält.

Analyse eines Spiels. Die Rahmenbedingungen autonomen Handelns werden als Spiel modelliert, um Vorhersagen über das Verhalten der Einheiten zu treffen. Dazu ist das Modell des Spiels zu analysieren. Im Folgenden werden die Methoden der Analyse am Beispiel des Gefangenendilemmas eingeführt.

Zu welcher Strategie werden sich die Ladenbesitzer entscheiden? Unabhängig davon, welche Strategie der Gegenspieler verfolgt, erhält ein Ladenbesitzer eine höhere Auszahlung, wenn er sich zu niedrigen Preisen entscheidet. Die Wahl niedriger Preise stellt daher eine *dominante* Strategie dar. Umgekehrt ist die Wahl eines hohen Preises eine *dominierte* Strategie. Es liegt daher nahe vorherzusagen, dass beide Ladenbesitzer sich zu niedrigen Preisen entscheiden werden. Allerdings müssen wir hierfür eine Annahme treffen, die die Grundlage der Betrachtungen der Spieltheorie bildet. Sie besagt, dass sich die Spieler in dem Sinne *rational* verhalten, dass sie diejenige Strategie

Tabelle 4.2: Beispiel für ein Koordinationsspiel

		Spieler B	
		Zahl	Wappen
Spieler A	Zahl	$(1, 1)$	$(0, 0)$
	Wappen	$(0, 0)$	$(1, 1)$

auswählen, die ihre erwartete Auszahlung maximiert. Eine Mindestanforderung der Rationalität ist also, dass kein Spieler eine dominierte Strategie verfolgt. Weiterhin wird in der Spieltheorie angenommen, dass die Rationalität der Spieler *gemeinsames Vorwissen* (engl.: common knowledge) darstellt. Dies bedeutet Folgendes:

- Jeder Spieler weiß, dass alle Spieler rational sind.
- Jeder Spieler weiß, dass alle Spieler wissen, dass alle Spieler rational sind.
- Für jede beliebige Schachtelungsstufe des Wissens gilt die Aussage weiterhin.

Aufgrund der Annahme der Rationalität können wir vorhersagen, dass beide Ladenbesitzer sich zu niedrigen Preisen entscheiden. Dieser Schluss erscheint auf den ersten Blick unintuitiv, da beide Ladenbesitzer bei einer beidseitig hohen Preissetzung eine höhere Auszahlung erhalten würden. Diese Intuition schlägt sich in der folgenden Definition nieder: Ein Ausgang ist *Pareto-effizient*, wenn es keinen anderen Ausgang gibt, bei dem alle Spieler mindestens eine ebenso hohe Auszahlung erhalten. Das beidseitige Wählen niedriger Preise ist also nicht Pareto-effizient, da die Auszahlungen der Ladenbesitzer jeweils geringer sind als bei beidseitig hohen Preisen. Allerdings sind auch die Ausgänge, bei denen die Ladenbesitzer unterschiedliche Preise festsetzen, Pareto-effizient. Dies liegt daran, dass der Ladenbesitzer, der den niedrigen Preis wählt, seine größtmögliche Auszahlung erhält und daher alle anderen Ausgänge, insbesondere den Ausgang beidseitig hoher Preise, weniger vorteilhaft findet. Eben dies ist der Grund, dass die Spieler sich zu niedrigen Preisen entscheiden.

Gleichgewichte. Die Analyse des Gefangenendilemmas wird durch die Tatsache erleichtert, dass beide Spieler eine dominante Strategie besitzen. Dass dies jedoch nicht bei jedem Spiel der Fall sein muss, zeigt das Koordinationsspiel aus Tabelle 4.2. In ihm treffen zwei Spieler aufeinander, die gleichzeitig je eine Münze hinlegen müssen. Entscheiden sich beide dafür, die Münze mit derselben Seite nach oben hinzulegen, so erhalten sie eine Auszahlung von 1. Ansonsten gehen die Spieler leer aus. Die bisher eingeführten Methoden der Analyse erlauben keine Vorhersage über die Wahl der Strategien der beiden Spieler. Dies liegt daran, dass weder das Auflegen von Zahl noch das von Wappen eine dominante Strategie darstellt.

Um dennoch zu Vorhersagen über das Verhalten der Spieler zu gelangen, benötigen wir ein mächtigeres Lösungskonzept. Zu diesem Zweck bietet die Spieltheorie das Konzept der Gleichgewichte an. Die Strategien der Spieler befinden sich in einem *Gleichgewicht*, wenn keiner der Spieler – wissend welche Strategie die anderen Spieler verfolgen – einen Anreiz hat, seine eigene Strategie zu ändern. Im Koordinationsspiel gibt es also zwei Gleichgewichte. Sie liegen in den beiden Strategiekombinationen, die jeweils zu einem der Pareto-effizienten Ausgänge *(Zahl, Zahl)* und *(Wappen, Wappen)* führen. Woher wissen aber die Spieler, welches der Gleichgewichte sie anstreben sollen? Dazu benötigen sie Informationen darüber, welche Strategie von ihrem jeweili-

gen Gegenspieler verfolgt werden wird. Eine solche Information ist den Spielern allerdings nicht zugänglich, wie aus der Modellierung des Koordinationsspiels als simultanes Spiel deutlich wird.

Eine Ergänzung des Strategiebegriffs führt zu einer Erweiterung des Gleichgewicht-Lösungskonzepts. Bisher sind wir davon ausgegangen, dass Strategien deterministisch festlegen, welche Handlungen in Abhängigkeit der Spielsituation erfolgen. Solche *reinen Strategien* stellen jedoch nicht die einzige Möglichkeit dar, nach der sich ein Spieler verhalten kann. Ebenso ist es denkbar, dass ein Spieler eine *gemischte Strategie* verfolgt, in der für jede Spielsituation festgelegt ist, mit welcher Wahrscheinlichkeit eine Handlung erfolgt. Im Koordinationsspiel stellt sich eine solche gemischte Strategie dadurch dar, dass ein Spieler mit der Wahrscheinlichkeit p die Zahl und mit $(1-p)$ das Wappen wählt. Gemischte Strategien stellen eine Verallgemeinerung von reinen Strategien dar. Zum Beispiel können die reinen Strategien des Koordinationsspiels mit $p = 0\%$ und $p = 100\%$ charakterisiert werden. Findet sich ein Gleichgewicht in gemischten Strategien, so nennen wir es *gemischtes Gleichgewicht*. Im Koordinationsspiel gibt es ein solches Gleichgewicht, wenn beide Spieler die Strategie mit $p = 50\%$ verfolgen.

Das Koordinationsspiel besitzt somit insgesamt drei Gleichgewichte. Wenn wir Vorhersagen über das Verhalten der Spieler treffen wollen, stellt sich damit die Frage, ob ein Gleichgewicht eintritt und welches es ist. Nützlich ist hierbei das Charakteristikum der Stabilität:

- *Stabiles Gleichgewicht:* Weicht einer der Spieler mit seiner Strategie unwesentlich vom Gleichgewicht ab, so behalten die anderen Spieler ihre Gleichgewichtsstrategien bei. Im Koordinationsspiel sind die beiden Gleichgewichte in reinen Strategien solche stabile Gleichgewichte. Dass dem so ist, zeigen wir am Beispiel des Gleichgewichts in *(Zahl, Zahl)*: Wählt Spieler A die Zahl nur mit einer Wahrscheinlichkeit von $p = 1 - \epsilon$, so ist es für Spieler B weiterhin von Vorteil, bei seiner Strategie zu bleiben, immer die Zahl zu wählen.

- *Labiles Gleichgewicht:* Durch eine Abweichung eines Spielers von seiner Gleichgewichtsstrategie erhalten die anderen Spieler einen Anreiz, auch ihre Strategien zu ändern. Im Koordinationsspiel ist das gemischte Gleichgewicht labil. Im gemischten Gleichgewicht erhalten beide Spieler 0.5 als erwartete Auszahlung. Wählt Spieler A mit einer Wahrscheinlichkeit von zum Beispiel $p = 51\%$ Zahl, so ändert sich die erwartete Auszahlung beider Spieler zwar nicht. Spieler B erhält allerdings den Anreiz, immer Zahl zu wählen. In diesem Fall würde nämlich die erwartete Auszahlung 0.51 betragen. Damit ist das gemischte Gleichgewicht labil.

Stabile Gleichgewichte sind demnach insofern eher zu erwarten, als sie sich aus Abweichungen aus labilen Gleichgewichten ergeben. Allerdings ist im Koordinationsspiel unklar, welches der beiden stabilen Gleichgewichte bei einem Spiel eintreten wird. Dies ist als Schwäche des Gleichgewichts als Lösungskonzept in der Literatur dargestellt worden.

4.1.2 Informationsasymmetrie und Signalisierung

Bei der Beschreibung der Grundzüge der Spieltheorie sind wir davon ausgegangen, dass die Spieler über dieselbe Information darüber verfügen, welche Nutzenstruktur jeder teilnehmender Spieler besitzt. Solche Spiele mit *symmetrischer Information* können besonders elegant behandelt werden, da für jeden Spieler bekannt ist, auf welcher Grundlage er handelt. Allerdings kann in realistischen Umgebungen wie auch der des Campus-Szenarios nicht davon ausgegangen werden, dass jeder Spieler stets über dieselben Informationen verfügt. Es stellt sich daher die Frage, wie Spiele mit solcher *asymmetrischen Information* zu behandeln sind.

In diesem Abschnitt befassen wir uns mit den Erweiterungen der Spieltheorie, die Antworten auf diese Frage geben. Dazu werden zunächst die Grundzüge der Informationsasymmetrie besprochen. Nachfolgend untersuchen wir, wie ein Spieler sein Wissen signalisieren kann. Abschließend gehen wir auf das Ladenketten-Paradox ein, das Elemente sowohl der Informationsasymmetrie als auch der Signalisierung beherbergt.

Die Ausführungen dieses Abschnitts beschränken sich auf die Vorstellung derjenigen Konzepte, die für die weiteren Betrachtungen dieser Arbeit absolut notwendig sind. Einen umfassenden Überblick über die Theorie der Spiele unter Informationsasymmetrie findet sich in RASMUSEN [Ras89].

Informationsasymmetrie. Sind gewisse Informationen nur einem Teil der Spieler bekannt, so sprechen wir von Informationsasymmetrie. Informationen können sich dabei auf unterschiedliche Bereiche des Spiels beziehen:

- *Nutzenstruktur:* Sie ist durch die Auszahlungen eines Spielers für unterschiedliche Spielergebnisse gegeben.
- *Handlungsrepertoire:* Das Repertoire bestimmt, zu welchen Handlungen ein Spieler fähig ist oder welche Handlungen aufgrund seiner Nutzenstruktur bei der Wahl seiner Strategie überhaupt berücksichtigt werden. Im Extremfall besitzt ein Spieler nur eine Handlungsmöglichkeit. Wir nennen dann diesen Spieler *vor-festgelegt* (engl.: pre-committed) zu gewissem Verhalten.
- *Informationen anderer:* Ein Spieler weiß nicht, über welche Informationen andere Spieler verfügen. Daher können sich Informationen auch darauf beziehen, welche Informationen welchem Spieler zugänglich sind.

Informationen zu diesen drei Bereichen beziehen sich jeweils auf die Eigenschaften eines Spielers. Wenn zwei Spieler dieselben Eigenschaften besitzen, so sprechen wir davon, dass sie denselben *Typ* besitzen [KS92].

Fehlen einem Spieler Informationen über einen dieser Bereiche, so muss er sich im Verlauf des Spiels darüber einen *Glauben bilden*. Nur dann kann er sein Verhalten an der jeweiligen Spielsituation ausrichten. Der Glauben ist immer dann zu revidieren, wenn der Spieler durch Beobachtung des Spielverlaufs neue Informationen erhält.

Diese Überlegungen verdeutlichen wir anhand eines Spiels, wie es in Tabelle 4.3 aus der Sicht eines Spielers X dargestellt ist. Beide Spieler können wählen, wie stark sie sich für die Erreichung eines gemeinsamen Zieles einsetzen. Für jeden Spieler bedeutet hoher Einsatz einen Aufwand, der die Auszahlung um 1 verringert. Die Spieler erreichen nur dann ihr gemeinsames Ziel, wenn sie beide sich für einen hohen Einsatz entscheiden. Für Spieler X bringt die Zielerreichung einen Nutzen von 2 mit sich. Allerdings hat er keine Informationen darüber, wie hoch der Nutzen α von Spieler Y bei Zielerreichung ist. Alleine Spieler Y verfügt über die Information, wie hoch α ist. Es handelt sich also um ein Spiel mit asymmetrischer Information. Weiterhin nehmen wir an, dass Spieler X aufgrund der Rahmenbedingungen des Spiels weiß, dass es genau zwei Typen von Spielern gibt: **(1)** Spieler vom Typ T_1 sind auf die Zielerreichung angewiesen. Für sie ist $\alpha = 2$. Diesem Typ gehört auch Spieler X an. **(2)** Hingegen erzielen Spieler vom Typ T_2 kaum einen Nutzen aus der Zielerreichung. Für sie ist $\alpha = 0$.

Wie muss das Spiel aus der Sicht von Spieler X analysiert werden? Wenn Spieler Y vom Typ T_1 ist, handelt es sich um ein Koordinationsspiel, bei dem beidseitiger hoher Einsatz aufgrund der Paretoeffizienz eines solchen Ausgangs zu erwarten ist [FT91]. Ist Spieler Y allerdings vom

Tabelle 4.3: Beispiel für ein Spiel unter Informationsasymmetrie

		Y	
		hoher Einsatz	niedriger Einsatz
X	hoher Einsatz	$(2, \alpha)$	$(-1, 0)$
	niedriger Einsatz	$(0, -1)$	$(0, 0)$

Typ T_2, so ist für ihn eine dominante Strategie, immer niedrigen Einsatz zu zeigen. Entscheidend für das Verhalten von Spieler X ist daher sein Glauben über den Typ von Spieler Y. Es stellt sich also die Frage, wie er einen solchen Glauben bilden kann. Eine Möglichkeit besteht darin, dass Spieler X das Verhalten von Spieler Y in früheren Spielen dieser Art beobachtet. Eine Einordnung der Beobachtungen durch die Revision des eigenen Glaubens erfolgt dann wie folgt:

- *Beobachtung hohen Einsatzes:* Hat Spieler X beobachtet, dass Spieler Y in einem früheren Spiel hohen Einsatz gezeigt hat, so ist er sich sicher, dass Spieler Y vom Typ T_1 ist. Wäre jener nämlich von Typ T_2, so würde er immer niedrigen Einsatz zeigen.

- *Beobachtung niedrigen Einsatzes:* Wird ein niedriger Einsatz von Spieler Y beobachtet, so kann Spieler X keine direkten Schlussfolgerungen für seine Glaubensrevision ziehen. Dies liegt daran, dass auch ein Spieler vom Typ T_1 sich zu niedrigem Einsatz entscheiden kann. Dies gilt insbesondere dann, wenn er erwartet, dass sein Gegenspieler vom Typ T_2 ist.

Anhand des Beispiels lässt sich also die Grundproblematik der Informationsasymmetrie zeigen: Der Mangel an Information erfordert eine Glaubensbildung, deren Durchführung wiederum die Verfügbarkeit weiterer Information (im Beispiel die Beobachtungen des Verhaltens in früheren Spielen) benötigt. Im nächsten Paragraphen wird daher ein Konzept der Spieltheorie vorgestellt, dass die Beschaffung von Informationen zwecks Glaubensbildung systematisiert.

Signalisierung und Screening. Ein wesentliches Charakteristikum von Spielen ist, dass das Verhalten eines Spielers auch Auswirkungen auf andere Spieler haben kann. Es ist daher für jeden Spieler unabdingbar, das Verhalten der anderen Spieler im Vorhinein einzuschätzen und basierend darauf über sein eigenes Verhalten zu entscheiden. Das Verhalten anderer Spieler wiederum hängt von ihrem jeweiligen Typ ab. Folglich ist es in Spielen mit asymmetrischer Information unabdingbar, Informationen über den Typ anderer Spieler zu erhalten. Die Methode, die hierfür im vorigen Beispiel zur Anwendung kam, besteht darin, das Verhalten eines Spielers zu beobachten und daraus Rückschlüsse auf seinen Typ zu ziehen. Um zu solchen Beobachtungen zu kommen, ist es unter Umständen erforderlich, den zu untersuchenden Spieler in eine Spielsituation zu bringen, in der die Wahl seines Verhaltens solche Rückschlüsse ermöglicht. Diese Methode wird *Screening* genannt.

Die Lage stellt sich anders dar, wenn wir die Sicht eines Spielers annehmen, dessen Typ anderen Spielern nicht bekannt ist. Wie wir bereits gesehen haben, hängt das Verhalten dieser Spieler von ihrem Glauben über seinen Typ ab. Es bietet sich daher für den Spieler, dessen Typ unbekannt ist, an, durch bestimmtes Verhalten den anderen Spielern anzudeuten, von welchem Typ er ist. Diese Methode wird *Signalisierung* genannt. Sie ist das Gegenstück zum Screening: Ein Spieler, der einem Screening unterzogen wird, signalisiert seinen Typ durch sein Verhalten. Umgekehrt wird Verhalten, durch das ein Spieler proaktiv seinen Typ signalisiert, von anderen Spielern beobachtet.

Die Schwierigkeit bei der Bewertung von Signalen besteht darin, dass ein Spieler auch signalisieren kann, von einem Typ zu sein, dem er tatsächlich nicht angehört. Dass eine solche gewollte *Signalstörung* (engl.: signal jamming) von Vorteil sein kann, zeigt das *Einstellungsspiel* aus [Spe74]. Dessen Ausgangslage ist wie folgt charakterisiert:

- Der erste Spieler ist jemand, der sich auf eine Stelle in einem Unternehmen bewirbt. Der Unternehmer, der über Einstellung des Bewerbers entscheidet, ist der zweite Spieler.
- Dem Bewerber bringt eine Einstellung einen Nutzen von 2.
- Wir unterscheiden zwei Typen von Bewerbern, je nach dem ob sie talentiert sind oder nicht. Stellt der Unternehmer einen talentierten Bewerber ein, so bringt ihm dies einen Nutzen von 1. Wird jedoch ein untalentierter Bewerber eingestellt, so kostet den Unternehmer die Einstellung und er erhält einen negativen Nutzen von -1.

Aus der Ausgangslage geht hervor, dass beide Typen von Bewerbern anstreben, eingestellt zu werden, der Unternehmer jedoch nur talentierte Bewerber einstellen möchte. Aus der Sicht des Unternehmers ist es daher wünschenswert, Informationen über den Typ der Bewerber zu erhalten. Nehmen wir an, dass sich untalentierte Bewerber (zum Beispiel durch entsprechendes Training) im Bewerbungsgespräch als talentiert darstellen können. In diesem Fall führt die Durchführung eines Screenings mittels Bewerbungsgespräch zu keinen neuen Einsichten für den Unternehmer. Wir sprechen hier von einer *Zusammenlegung* (engl.: pooling), da die Bewerber unabhängig von ihrem Typ auf dieselbe Weise signalisieren. Die Ursache dafür ist, dass die Fähigkeit zur Signalisierung des eigenen Talents nicht vom Typ des jeweiligen Bewerbers abhängt. Das Signalisieren selbst verkommt daher zu *billigem Gerede* (engl.: cheap talk).

Um dennoch talentierte von untalentierten Bewerbern unterscheiden zu können, bedarf es einer anderen Signalisierungsmöglichkeit. In [BG97] werden zwei Bedingungen dafür gegeben, dass ein Signal *glaubhaft* ist und damit zu einer *Trennung* zwischen Spielern unterschiedlichen Typs führt. Diese stellen sich im Einstellungsspiel wie folgt dar:

- *Kann-Bedingung:* Für einen talentierten Bewerber ist es von Vorteil, seinen Typ wahrheitsgemäß zu signalisieren.
- *Kann-nicht-Bedingung:* Ein untalentierter Bewerber ist zwar in der Lage, ein Signal dafür zu geben, dass er talentiert sei. Ein solches nicht wahrheitsgemäßes Signalisieren verursacht ihm jedoch Kosten, die den Vorteil, vom Unternehmer als talentiert eingestuft zu werden, zunichte machen.

Beide Bedingungen können nur dann eingehalten werden, wenn die Art des Screenings von den Typ-abhängigen Eigenschaften der Bewerber Gebrauch macht. Im Einstellungsspiel dient hierfür die Option, vor einer Bewerbung eine Ausbildung zu absolvieren. Das Absolvieren der Ausbildung fällt talentierten Bewerbern leichter als untalentierten. Wir halten dies fest, indem die Kosten, die bei einer Ausbildung anfallen, für talentierte Bewerber auf 1 und für untalentierte Bewerber auf 3 angesetzt werden.

Tabelle 4.4 fasst das sich ergebende Einstellungsspiel zusammen. Die Auszahlungen des Unternehmers sind unterschiedlich, je nachdem ob ihm ein talentierter oder untalentierter Bewerber gegenübersteht. Dabei ist zu beachten, dass der Bewerber zuerst zieht. Das bedeutet, dass der Unternehmer für seine Einstellungsentscheidung berücksichtigen kann, ob der Bewerber eine Ausbildung gemacht hat. Die wesentlichen Punkte der Analyse des Spiels sind wie folgt:

Tabelle 4.4: Signalisierung im Einstellungsspiel

		Talentierter Bewerber	
		Ausbildung	keine Ausbildung
Unternehmer	Einstellung	$(1, 1)$	$(1, 2)$
	keine Einstellung	$(0, -1)$	$(0, 0)$

		Untalentierter Bewerber	
		Ausbildung	keine Ausbildung
Unternehmer	Einstellung	$(-1, -1)$	$(-1, 2)$
	keine Einstellung	$(0, -3)$	$(0, 0)$

- Für einen untalentierten Bewerber besteht die dominante Strategie darin, keine Ausbildung zu machen. Dies liegt daran, dass die zugehörigen Auszahlungen diejenigen im Falle einer Ausbildung übersteigen.

- Für einen talentierten Bewerber stellt sich die Situation anders dar. Er bevorzugt das Ergebnis *(Einstellung, Ausbildung)* gegenüber *(keine Einstellung, keine Ausbildung)*. Wenn der Unternehmer seine Einstellungsentscheidung von der Absolvierung einer Ausbildung abhängig macht, ist es für talentierte Bewerber von Vorteil, eine Ausbildung zu absolvieren[2].

Das Absolvieren einer Ausbildung stellt damit ein glaubhaftes Signal dafür dar, dass man ein talentierter Bewerber ist. Der Unternehmer wird diese Signalisierungsmöglichkeit ausnutzen, um nur talentierte Bewerber einzustellen. Dieser Schluss ist umso bemerkenswerter, als das Einstellungsspiel davon ausgeht, dass das Absolvieren einer Ausbildung nur Kosten aber keinen Nutzen für den Bewerber oder seinen Arbeitgeber mit sich bringt. Es liegt also in der Natur der Informationsasymmetrie, dass durch sie an sich nachteilige Handlungen (im Einstellungsspiel das Absolvieren einer Ausbildung) gewählt werden.

Aus dem Einstellungsspiel wird deutlich, dass die Methode der Signalisierung sorgsam gewählt werden muss. Nur dann ist es möglich, zu verlässlichen Informationen über den Typ anderer Spieler zu gelangen und darauf aufbauend die eigenen Handlungen entsprechend zu wählen.

Informationsasymmetrie und Signalisierung im Ladenkettenspiel. Für die weiteren Betrachtungen dieser Arbeit ist das Ladenkettenspiel (engl.: chain-store game) [Sel78] von Bedeutung. Es ist ein weiteres Beispiel dafür, dass an sich nachteilige Handlungsvarianten bei asymmetrischer Verteilung von Information vorteilhaft zum Zwecke der Signalisierung eingesetzt werden können. Das Ladenkettenspiel gestaltet sich wie folgt:

- Bei den Spielern handelt es sich um einen Monopolisten und einen potentiellen Konkurrenten, der die Wahl hat in den Markt einzutreten. Kommt der Markteintritt zustande, so kann sich der Monopolist zu einem Preiskampf oder zu einer Marktaufteilung mit dem Konkurrenten entscheiden.

[2] Würde der Bewerber nicht vor dem Unternehmer ziehen, so würde auch für talentierte Bewerber die dominante Strategie darin bestehen, keine Ausbildung zu absolvieren. Den hier vorgebrachten Schluss erhalten wir also nur wegen der Sequentialität der Handlungen der Spieler.

Tabelle 4.5: Signalisierung im Ladenkettenspiel

		Schwacher Monopolist	
		Preiskampf	Marktaufteilung
Konkurrent	Eintritt	$(-2, -2)$	$(1, -1)$
	kein Eintritt	$(0, 0)$	$(0, 0)$

		Starker Monopolist	
		Preiskampf	Marktaufteilung
Konkurrent	Eintritt	$(-2, -1)$	$(1, -2)$
	kein Eintritt	$(0, 0)$	$(0, 0)$

- Als Ausgang des Spiels bevorzugt der potentielle Konkurrent den Markteintritt mit nachfolgender Marktaufteilung. Sich einem Preiskampf auszusetzen, stellt für ihn den schlechtesten Ausgang dar.

- Der Monopolist wünscht sich, dass es erst gar nicht zum Markteintritt des Konkurrenten kommt. Auch für ihn ist ein Preiskampf der schlechtest mögliche Ausgang.

Wie werden sich der potentielle Konkurrent und der Monopolist aufgrund dieser Ausgangslage verhalten? Kommt es zum Markteintritt, so wird der Monopolist sich zur Marktaufteilung entschließen, um einen Preiskampf zu vermeiden. Dies weiß auch der potentielle Konkurrent. Für ihn besteht damit die dominante Strategie darin, in den Markt einzutreten.

Das Ergebnis dieser Analyse widerspricht den empirischen Befunden über das Verhalten von Kettenläden [Sel78]. Dort wurde beobachtet, dass der Markteintritt häufig mit Preiskampf beantwortet wird. Diesen Widerspruch zwischen spieltheoretischer Vorhersage von rationalem Verhalten und dem empirischen Befund wird in der Literatur als *Ladenketten-Paradox* bezeichnet.

Zur Lösung dieses Paradoxons wird in [KW82] die Informationsasymmetrie herangezogen. Sie besteht darin, dass dem potentiellen Konkurrenten Informationen über die Präferenzen des Monopolisten fehlen. Es gibt zwei Typen von Monopolisten: Der erste Typ entspricht der bisherigen Beschreibung des Monopolisten. Es handelt sich um einen schwachen Monopolisten, da er den Preiskampf mehr fürchtet als eine Marktaufteilung. Im Gegensatz dazu stellt für einen starken Monopolisten eine Marktaufteilung keine Option dar. Er ist daher auf Preiskämpfe mit Konkurrenten vor-festgelegt. Die Präferenzen der Spieler werden in Tabelle 4.5 mit beispielhaften Zahlenwerten zusammengefasst.

Wie wird ein potentieller Konkurrent seiner Unsicherheit über den Typ des Monopolisten begegnen? Handelt es sich um einen schwachen Monopolisten, so lohnt sich der Markteintritt, da eine Marktaufteilung zu erwarten ist. Ist der Monopolist allerdings stark, so sollte der potentielle Konkurrent von einem Markteintritt absehen, um einen Preiskampf zu vermeiden. Da der potentielle Konkurrent also sein eigenes Verhalten gerne vom Typ des Monopolisten abhängig machen würde, ist ein Screening über den Monopolisten durchzuführen. Im Ladenkettenspiel geht dies nur über Beobachtungen darüber, wie der Monopolist frühere Markteintritte beantwortet hat. Wenn er sich jemals zur Marktaufteilung entschieden hat, dann handelt es sich um einen schwachen Monopolisten. In diesem Fall ist ein Markteintritt für den potentiellen Konkurrenten lohnenswert.

Diese Überlegungen des potentiellen Konkurrenten sind auch dem Monopolisten bekannt. Für einen schwachen Monopolisten ergibt sich daraus der folgende Schluss: Er muss Markteintritte

mit Preiskämpfen beantworten, damit es nicht zu weiteren Markteintritten anderer Konkurrenten kommt. Der Preiskampf stellt für den schwachen Monopolisten daher eine Möglichkeit zur Signalisierung seiner vermeintlichen Stärke dar. Dabei nimmt der Monopolist die Kosten des Preiskampfes hin, um ein solches Signal auszusenden. Das Signal ist also umso glaubwürdiger, je größer die Kosten des Preiskampfes für den schwachen Monopolisten sind. Dieser Sachverhalt wird in [KW82] derart beschrieben, dass der Monopolist sich kurzfristig gesehen irrational verhält, um sich einen längerfristigen Nutzen zu sichern.

Das Ladenkettenspiel unter Informationsasymmetrie findet eine Erweiterung in [MR82]. Dort wird angenommen, dass der potentielle Konkurrent den Typ des Monopolisten kennt. Allerdings ist sich der Monopolist nicht sicher, dass sein Typ vom potentiellen Konkurrenten erkannt wird. Die vorige Analyse kann für diese Abschwächung der Informationsasymmetrie analog durchgeführt werden. Auch wenn der potentielle Konkurrent weiß, dass es sich um einen schwachen Monopolisten handelt, wird er unter Umständen nicht in den Markt eintreten. Dies liegt daran, dass der schwache Monopolist sich weiterhin zur Signalisierung seiner vermeintlichen Stärke entscheiden könnte. Eine weitere Analyse zeigt, dass wir dieses Ergebnis immer dann erhalten, wenn der Typ des Monopolisten kein gemeinsames Wissen (engl.: common knowledge) ist. Wenn zum Beispiel der Monopolist weiß, dass der potentielle Konkurrent seinen Typ kennt, sich aber unsicher ist, ob der Konkurrent von diesem Wissen Kenntnis hat, so wird sich auch ein schwacher Monopolist weiterhin zu Preiskämpfen entscheiden. Diesen Zusammenhang zwischen gemeinsamen Wissen und Informationsasymmetrie werden wir in Abschnitt 7.6.2 ausnutzen.

4.1.3 Evolutionäre Spieltheorie

Die bisher vorgestellten Konzepte der Spieltheorie befassen sich damit, für gegebene Rahmenbedingungen vorherzusagen, wie sich rationale Einheiten verhalten. Die evolutionäre Spieltheorie geht einen anderen Weg: Ausgehend von einer Zuordnung von Strategien zu Einheiten wird untersucht, welche Strategien sich bei mehrmaliger Kooperation zwischen den Einheiten untereinander durchsetzen.

Dieser Abschnitt stellt die Grundzüge der evolutionären Spieltheorie vor und diskutiert die Erweiterungen, die für sie in der Literatur vorgeschlagen worden sind. Es zeigt sich, dass ihr Modell der Kooperation in mancher Hinsicht äquivalent ist mit dem Systemmodell von Informationssystemen wie im Campus-Szenario. Abschließend wird die Kritik vorgetragen, die in der Literatur an der evolutionären Spieltheorie geübt worden ist.

Grundzüge. Eine vielzitierte Arbeit zur evolutionären Spieltheorie ist *Axelrods Kooperationsturnier* [Axe84]. Es wird zwischen einer Menge von Einheiten ausgetragen, die jeweils eine ganz spezielle Strategie verfolgen. Der Ablauf des Turniers ist wie folgt: Es wird zufällig ein Paar von Einheiten ausgewählt, die als Spieler im iterierten Gefangenendilemma (engl.: iterated prisoners' dilemma), kurz *IPD*, gegeneinander antreten. Das *IPD* erweitert das Gefangenendilemma darin, dass beide Einheiten an einer Reihe von Gefangenendilemma-Spielen teilnehmen. Dadurch besteht die Strategie einer Einheit nicht nur aus der Auswahl der Handlungsalternative für die erste Iteration. Darüber hinaus definiert sie, wie auf das Verhalten des Gegenübers in nachfolgenden Iterationen reagiert wird. Beispiel für eine solche Strategie ist *Tit-for-Tat* (TFT), das in der ersten Iteration kooperatives Verhalten vorschreibt und danach das Verhalten des Gegenübers aus der jeweils vorigen Iteration wiederholt. Damit beantwortet *TFT* Kooperation mit Kooperation und Betrug mit Betrug. Einfachere Strategien sind *immer kooperativ* (engl.: all cooperative, kurz *allC*) und *immer betrügend* (engl.: all defective, kurz *allD*), die in jeder Iteration kooperatives

beziehungsweise betrügendes Verhalten vorsehen. Der Individualnutzen der einzelnen Einheiten, der sich aus ihrer Teilnahme an verschiedenen *IPD*-Spielen ergeben hat, wird dabei festgehalten. Dieser Individualnutzen wird im evolutionären Kontext *Fitness* genannt. Er entscheidet darüber, welche Strategien die Einheiten in der nächsten Turnierrunde verfolgen. Diejenigen Strategien, die von Einheiten mit hoher Fitness gespielt wurden, sind dabei besonders stark vertreten. In dieser Hinsicht findet eine Selektion unter den Strategien statt. Der Zusammenhang zwischen dem Erfolg einer Strategie und ihrer anteilsmäßigen Verfolgung durch die Einheiten führt zu einer Dynamik, die *Replikationsdynamik* genannt wird. Je nach Einstellung des Turniers werden die selektierten Strategien zudem leicht abgeändert (Mutation) oder miteinander verbunden (Rekombination).

Mit dem hier beschriebenen Kooperationsturnier verfolgt Axelrod eine Methodik der wissenschaftlichen Erkenntnisbildung, die im Gegensatz zur klassischen Spieltheorie steht. Jene setzt analytische Verfahren ein, um ausgehend von den Modelleigenschaften auf deduktive Weise zu Aussagen zu kommen. Die Durchführung des Kooperationsturniers hingegen erfolgt automatisiert im Rahmen einer computer-gestützten *Simulation*. Axelrod verteidigt diese simulative Methodik in [Axe97]. Ihre Anwendung ist dann von Vorteil, wenn das Kooperationsmodell zu komplex ist oder die Einheiten ihr Verhalten adaptiv aneinander ausrichten.

Die Durchführung von Axelrods Kooperationsturnier zeigt, dass sich *TFT* gegen alle anderen eingereichten Strategien durchsetzt. Axelrod folgert daraus, dass *TFT* eine *evolutionär stabile Strategie* darstellt. Dieses Stabilitätskonzept wird in [MS82] eingeführt und besagt, dass eine Strategie – wenn sie einmal von allen Einheiten verfolgt wird – nicht mehr durch eine in geringer Zahl eindringende andere Strategie verdrängt werden kann. Axelrod basiert seine Schlussfolgerung darauf, dass eine Einheit, die *TFT* verfolgt, **(1)** nie als erstes betrügt und damit den Kooperationsvorteil mit anderen kooperationswilligen Einheiten voll ausnutzt, **(2)** sie immer nur unwesentlich schlechter als betrügende Einheiten abschneidet, da sie Betrug mit Betrug beantwortet und **(3)** sie auf erneute Kooperationsangebote eines Gegenübers, der zuvor betrogen hat, sofort eingeht.

Das Konzept evolutionär stabiler Strategien gibt keinen Aufschluss darüber, wie sich die Menge der verfolgten Strategien aufgrund der Replikationsdynamik entwickelt. Hiermit beschäftigt sich die Theorie der *Selbstorganisation*, wie sie in [Kau93] mit der Theorie der Evolution verbunden worden ist. Die Replikationsdynamik entsteht aus der *Rückkopplung* zwischen der Selektion der Strategien einerseits und dem im Turnier vorherrschenden Verhalten andererseits. Diese Rückkopplung ist in Abbildung 4.1 dargestellt. Sie bewirkt, dass sich die *Populationsstruktur*, das heißt die Zuteilung der Strategien zu den Einheiten, dynamisch entwickelt. Kommt diese dynamische Entwicklung in einer Populationsstruktur zum Stehen, so wird diese *Attraktor* genannt [Hey04]. Zum Beispiel wird in Axelrods Turnier eine Konvergenz zu einer Populationsstruktur beobachtet, in der nur *TFT* vertreten ist. In diesem Fall stellt diese reine *TFT*-Population einen Attraktor dar.

Erweiterungen. Das Modell der Kooperation ist im iterierten Gefangenendilemma zu speziell, als dass es für allgemeine Kooperationsumgebungen zutreffen würde. In der Literatur sind daher einige Erweiterungen zum *IPD* aus Axelrods Turnier eingeführt worden. Eine Übersicht hierzu findet sich in [Axe00]. Im Folgenden werden die Erweiterungen vorgestellt, durch die das Kooperationsmodell des *IPD* an das des Campus-Szenario angeglichen wird:

- *Wiedererkennbarkeit der Einheiten:* Die Einheiten können untereinander ihre jeweilige *Identität* erkennen. Das bedeutet, dass Einheiten, die in einem *IPD* aufeinander getroffen sind, sich wieder erkennen, wenn sie in an einem weiteren *IPD* teilnehmen. Diese Erweiterung

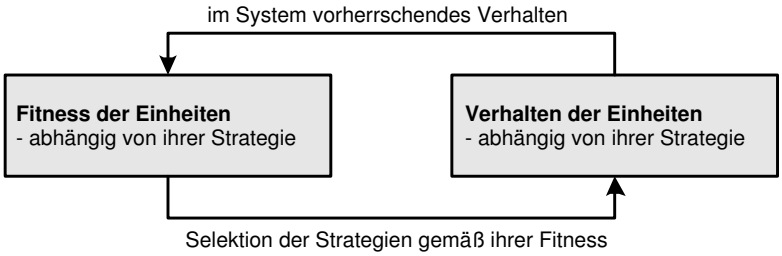

Abbildung 4.1: Rückkopplung als Ursache der Replikationsdynamik

des Kooperationsmodells ist die Voraussetzung dafür, dass sich Einheiten eine Reputation aufbauen können [CP02].

- *Lokalisierung der Kooperation:* Bei der Gestaltung der Simulationsumgebung stellt sich die Frage, welches Paar von Einheiten jeweils an einem *IPD* teilnimmt. Axelrods Turnier wählt hierfür die Einheiten zufällig aus. Alternativ dazu wird etwa in [CCP98] vorgeschlagen, dass jede Einheit eine Lokation zugewiesen bekommt und lokal benachbarte Einheiten von der Simulationsumgebung als Paare eines *IPD* gewählt werden. Diese Lokalisierung der Kooperation entspricht der Situation im Campus-Szenario, da dort aufgrund der Partitionierung des Ad-hoc Netzes Kooperation stets lokalisiert stattfindet. Die Erweiterung des Kooperationsmodells um Lokalität hat zur Folge, dass sich die Folgekosten von Betrugsverhalten erhöhen.

- *Wahl des Gegenübers:* In [SSA94, ASST96] wird das iterierte Gefangenendilemma mit Wahl und Ablehnung (engl.: iterated prisoners' dilemma with choice and refusal), kurz *IPD/CR*, als Erweiterung des reinen *IPD* vorgestellt. Die Einheiten können wählen, mit welchen anderen Einheiten sie in ein *IPD*-Spiel treten. Insbesondere sind die Einheiten in der Lage, ein *IPD*-Spiel mit einer unerwünschten Einheit abzulehnen. Die Simulationsergebnisse zeigen, dass sich durch diese Möglichkeit zum Wählen und Ablehnen die Folgekosten von Betrug erhöhen [SAS95]. Das *IPD/CR* entspricht genau unserem Transaktionsmodell aus Abschnitt 1.2.2, in dem die Einheiten zwischen ihren potentiellen Transaktionspartnern wählen können.

- *Gestörte Wahrnehmung:* In Axelrods Turnier nehmen die Einheiten das Verhalten ihres jeweiligen Gegenübers korrekt wahr. Nimmt eine Einheit zum Beispiel Betrugsverhalten wahr, so ist sie sich sicher, dass der Gegenüber den Betrug beabsichtigt hat. Diese Situation widerspricht dem Campus-Szenario, da dort unbeabsichtigtes Fehlverhalten möglich ist. Eine Erweiterung des *IPD* muss daher dafür sorgen, dass die Wahrnehmung von Verhalten gestört ist. Eine solche Störung wird Rauschen (engl.: noise) genannt. Arbeiten wie [WA95] gehen daher der Frage nach, wie *TFT* für diesen Fall der gestörten Wahrnehmung angepasst werden muss. Die erfolgreichste Anpassung ist das *großzügige TFT* (engl.: generous TFT), kurz *GTFT*. Es berücksichtigt die Wahrscheinlichkeit für eine inkorrekte Wahrnehmung, um die Reaktion auf betrügendes Verhalten festzulegen.

Diese vier Erweiterungen von Axelrods Kooperationsturnier bilden eine Voraussetzung dafür, dass etwaige Simulationsergebnisse relevant für das Campus-Szenario sind. In der Literatur werden allerdings in keiner Studie all diese Erweiterungen berücksichtigt. Am Ehesten kommt die

Simulationsumgebung aus [SAS95] dem Kooperationsmodell des Campus-Szenarios nahe. Ihm fehlt jedoch die Berücksichtigung von lokalisierter Kooperation und von gestörter Wahrnehmung.

Kritik an der evolutionären Spieltheorie. In der Literatur ist von mehreren Seiten die Relevanz der evolutionären Spieltheorie zumindest teilweise in Frage gestellt worden. Einen Hauptangriffspunkt bietet das Kooperationsmodell, das den Betrachtungen der evolutionären Spieltheorie zugrunde gelegt wird. Eben aus diesem Grund ist es zu den bereits besprochenen Erweiterungen des Kooperationsmodells gekommen. Darüber hinaus gibt es jedoch weitere Kritikpunkte, die durch solche Erweiterungen nicht beseitigt werden können. Sie werden im Folgenden diskutiert.

Die Arbeit [Cha98] fasst einige dieser Kritikpunkte zusammen. Die Replikationsdynamik spiegelt zwar biologische Vererbungsprozesse wider. Sie ist aber ungeeignet, soziale Prozesse adäquat darzustellen. Im Campus-Szenario wählt zum Beispiel ein Benutzer, welche Version der Systemsoftware er verwenden will, und nicht umgekehrt Versionen der Systemsoftware, von welchen Benutzern sie verwendet werden wollen[3].

Ein weiterer Kritikpunkt betrifft das Konzept der evolutionär stabilen Strategien. Es macht nämlich keinerlei Aussagen darüber, welche Populationsstruktur ein Attraktor ist. Eine Strategie kann das Kriterium der evolutionären Stabilität erfüllen, ohne dass eine reine Populationsstruktur aus dieser Strategie einen Attraktor darstellt. Selbst wenn zum Beispiel die Strategie *TFT* evolutionär stabil ist, muss es aus einer beliebigen Ausgangssituation nicht notwendigerweise dazu kommen, dass alle Einheiten diese Strategie verfolgen. Die Definition von evolutionärer Stabilität fordert nämlich nur, dass eine Strategie, nachdem sie sich durchgesetzt hat, nicht mehr verdrängt werden kann. Ob es zu einem solchen Durchsetzen kommt, bleibt dabei allerdings unklar. Damit ist die Aussagekraft des Konzeptes der evolutionär stabilen Strategien beschränkt.

Hinzu kommt, dass die Eigenschaft der evolutionären Stabilität von den Rahmenbedingungen der Kooperation abhängt [Hof00]. Bei diesen Rahmenbedingungen ist vor allem die Zahl der Iterationen im *IPD* oder das Verhältnis zwischen dem Kooperationsvorteil und den Kosten des Betrogenwerdens entscheidend. Ist zum Beispiel die Zahl der zu erwartenden Iterationen zu niedrig oder der Betrugsvorteil besonders hoch, so schneidet die Strategie *allD* besser als *TFT* ab. In diesen Rahmenbedingungen kann damit *TFT* nicht evolutionär stabil sein. Aussagen über Strategien sind daher immer von den Rahmenbedingungen der Kooperation abhängig.

Auch die Schlussfolgerungen aus Axelrods Kooperationsturnier sind angegriffen worden. Unter anderem ist gezeigt worden, dass *TFT* nicht evolutionär stabil ist [Hof00]. Dies liegt daran, dass zum Beispiel die Strategie *allC* gegen *TFT* genauso gut abschneidet wie *TFT* untereinander. Als Folge davon ist die Selektion der Strategien indifferent zwischen *allC* und *TFT*, so dass die Strategie *TFT* die Strategie *allC* nicht verdrängen kann. In der Gegenwart von *allC* schneidet jedoch die Strategie *allD* weitaus besser als *TFT* ab. Tritt diese Strategie ein, so kann es daher zu einer Verdrängung von *TFT* kommen. Damit ist die Strategie *TFT* nicht evolutionär stabil.

4.2 Weitere Grundlagen

Außer der Spieltheorie besitzt diese Arbeit einige weitere Grundlagen. Sie stammen aus den Bereichen Epistemologie, Stochastik und der Theorie verteilter Systeme. Die Besprechung dieser Grundlagen erfolgt in diesem Abschnitt. Sie ist bei weitem nicht so umfangreich wie die vorige Darstellung der Spieltheorie.

[3] Dieser Gegensatz wird in Abschnitt 5.3.4 genauer herausgestellt werden.

Glaubensbildung unter Unsicherheit. Bei der Besprechung von Unsicherheit und Glaubensbildung gehen wir von folgendem Leitbeispiel aus: Anna besitzt zwei Würfel, von denen einer sechs Seiten und der andere vier Seiten besitzt. Einen ihrer Würfel wird Anna spielen. Bob würde gerne einschätzen, welche Augenzahl sich beim Würfeln ergibt. Allerdings besitzt er keine Information darüber, mit welchem Würfel Anna spielen wird.

Im Beispiel ist sich Bob über das Ergebnis des Würfelns unsicher. Eine solche *Unsicherheit* besitzt zwei Komponenten [Hel97]:

- *Epistemische Unsicherheit:* Bob hat keine Information darüber, welcher Würfel zum Einsatz kommt. Die Unsicherheit, die dadurch entsteht, ist insofern epistemisch, als sie durch das Einholen von Information über Annas Wahl des Würfels eliminiert werden kann. Diese Art der Unsicherheit entspricht der, die bei den Spielen unter Informationsasymmetrie aus 4.1.2 vorherrscht: Nur Anna weiß, welchen Würfel sie wählen wird.
- *Stochastische Unsicherheit:* Selbst wenn Bob bekannt ist, mit welchem Würfel gespielt wird, so ist er sich über das Ergebnis des Würfelns unsicher. Dies liegt an der Zufälligkeit, mit der das Ergebnis bestimmt wird. Diese zweite Art der Unsicherheit wird stochastisch genannt.

Wie kann Bob die Wahrscheinlichkeit für ein bestimmtes Ergebnis des Würfelns trotz seiner Unsicherheit einschätzen? Eine solche Einschätzung erfordert, dass Bob einen Glauben darüber besitzt, mit welchem Würfel Anna spielen wird. Dieser Glauben besteht darin, dass Bob der Wahl eines Würfels eine *subjektive Wahrscheinlichkeit* zuordnet [Ber85]. In der Literatur wird ein solcher Glauben daher auch *probabilistisch* genannt [Bac90]. Wie bereits in Abschnitt 4.1.2 für den Umgang mit Informationsasymmetrie erwähnt wurde, ist ein solcher Glauben immer dann zu revidieren, wenn neue Informationen über den Sachverhalt zur Verfügung stehen.

Wenn zum Beispiel Bob anfangs keine Information über die Wahl der Würfel besitzt, so könnte er beiden Würfeln dieselbe Wahrscheinlichkeit zuordnen. Wir bringen dies zum Ausdruck, indem wir $p_{Bob}(6W) = 50\%$ schreiben, wobei $6W$ für die Wahl des Sechserwürfels steht. Der Index *Bob* gibt in dieser Notation an, dass es sich um Bobs subjektive Wahrscheinlichkeit handelt. Erfährt Bob davon, dass Anna mit dem Sechserwürfel spielen wird, so revidiert er seinen Glauben, indem er $p_{Bob}(6W) = 100\%$ setzt. Die Glaubensbildung über den Viererwürfel erfolgt analog. Sie wird im Folgenden nicht weiter betrachtet, da die Wahl des Viererwürfels $4W$ den komplementären Sachverhalt zu $6W$ (Notation hierfür $\overline{6W}$) darstellt, so dass jederzeit $p_{Bob}(4W) = 1 - p_{Bob}(6W)$ gilt.

Unter Zuhilfenahme des Glaubens gestaltet sich die Kombination von epistemischer und stochastischer Unsicherheit in eine Einschätzung zukünftiger Ereignisse wie folgt: Bob möchte einschätzen, mit welcher Wahrscheinlichkeit sich die Augenzahl Z ergibt. Die stochastische Unsicherheit drückt sich in den bedingten Wahrscheinlichkeiten $p(Z|6W)$ und $p(Z|4W)$ aus. Diese werden mit Bobs Glauben zu folgender Einschätzung kombiniert:

$$p_{Bob}(Z) = p_{Bob}(6W) \cdot p(Z|6W) + p_{Bob}(4W) \cdot p(Z|4W) \tag{4.1}$$

Der Index *Bob* der probabilistischen Einschätzung gibt dabei an, dass die Einschätzung von Bobs Glauben abhängt.

Bayes-Formel und Odds-Darstellung. Wir erweitern das Leitbeispiel dadurch, dass Anna zweimal würfelt. Dabei muss sie denselben Würfel benutzen. Bob ist also in der Lage, das Ergebnis des ersten Würfelns zu beobachten, bevor er einschätzt, welche Augenzahl sich beim zweiten Würfeln ergibt. Die *Bayes-Formel* gibt an, wie Bob auf probabilistisch fundierte Weise eine solche

Tabelle 4.6: Beispiele einiger Odds-Darstellungen von Wahrscheinlichkeiten

p	0	0.25	0.4	0.5	0.6	0.75	1
$\hat{p}$	∞	$\frac{3}{1}$	$\frac{3}{2}$	1	$\frac{2}{3}$	$\frac{1}{3}$	0

Beobachtung in eine Revision seines Glaubens einbezieht [Ber85]. Kommt es beim ersten Würfeln zur Zahl Z, so revidiert Bob seinen Glauben wie folgt:

$$p_{Bob}(6W|Z) = \frac{p(Z|6W) \cdot p_{Bob}(6W)}{p_{Bob}(Z)} \tag{4.2}$$

Dabei stellen in der Formel die Größen $p_{Bob}(6W)$ Bobs Glauben vor und $p_{Bob}(6W|Z)$ seinen Glauben nach seiner Beobachtung dar. Sie werden als *priorer* beziehungsweise *posteriorer* Glauben bezeichnet. Besitzt Bob beispielsweise zu Anfang einen Glauben von $p_{Bob}(6W) = 50\%$, so ergeben sich in Abhängigkeit des Ergebnis Z des ersten Würfelns:

- $Z \leq 4$: Aufgrund der Einschätzung $p_{Bob}(Z) = \frac{5}{24}$ ergibt sich $p_{Bob}(6W) = 40\%$.
- $Z > 4$: Wegen $p(Z|4W) = 0$ erhält Bob $p_{Bob}(6W) = 100\%$.

Die Berechnung der Glaubensrevision wird vereinfacht, wenn die subjektive Wahrscheinlichkeit in der *Odds-Darstellung* angegeben wird. Eine Umrechnung zwischen einer Wahrscheinlichkeit p und ihrer Odds-Darstellung $\hat{p}$ lässt sich mit Hilfe der Transformationsfunktion o wie folgt durchführen:

$$\hat{p} = o(p) = \frac{\overline{p}}{p} = \frac{1-p}{p} = \frac{1}{p} - 1 \quad , \quad p = o^{-1}(\hat{p}) = \frac{1}{1+\hat{p}} \tag{4.3}$$

Anschaulich gesprochen gibt die Odds-Darstellung $\hat{p}$ also das Verhältnis der Wahrscheinlichkeiten dafür an, dass ein Ereignis eintritt oder nicht. Ein Verhältnis von drei-zu-eins besagt zum Beispiel, dass der Eintritt des Ereignisses nur in einem von vier Fällen zu erwarten ist. Tabelle 4.6 gibt Beispiele für die Odds-Darstellung einiger Wahrscheinlichkeiten an.

Der Nutzen der Odds-Darstellung wird deutlich, wenn wir die Bayes-Formel auf die Revision einer Wahrscheinlichkeit in Odds-Darstellung anwenden. Der Übersichtlichkeit wegen wählen wir dazu die allgemeine Form des Bayes-Formel mit den Ereignissen A und B:

$$\hat{p}(A|B) = \frac{p(\overline{A}|B)}{p(A|B)} = \frac{\frac{p(B|\overline{A}) \cdot p(\overline{A})}{p(B)}}{\frac{p(B|A) \cdot p(A)}{p(B)}} = \frac{p(B|\overline{A}) \cdot p(\overline{A})}{p(B|A) \cdot p(A)} = \frac{p(\overline{A})}{p(A)} \cdot \frac{p(B|\overline{A})}{p(B|A)} = \hat{p}(A) \cdot \frac{p(B|\overline{A})}{p(B|A)} \tag{4.4}$$

In der Odds-Darstellung erfolgt eine Revision demnach durch ein einfaches Multiplizieren mit dem Verhältnis der bedingten Wahrscheinlichkeiten $\frac{p(B|\overline{A})}{p(B|A)}$. Dieses Verhältnis nennen wir im Folgenden *Revisionsfaktor*. Im Leitbeispiel beträgt der Revisionsfaktor abhängig von der Augenzahl Z:

- $Z \leq 4 \quad \Rightarrow \quad \frac{p(Z|4W)}{p(Z|6W)} = \frac{1}{4} \cdot (\frac{1}{6})^{-1} = \frac{3}{2}$
- $Z > 4 \quad \Rightarrow \quad \frac{p(Z|4W)}{p(Z|6W)} = 0 \cdot (\frac{1}{6})^{-1} = 0$

Im Leitbeispiel vollzieht sich damit die Glaubensrevision wie folgt: Bob besitzt initial den Glauben von $p_{Bob}(6W) = 50\%$, den er in der Odds-Darstellung mit $\hat{p}_{Bob}(6W) = o(50\%) = 1$ überführt. Beobachtet Bob das Ergebnis $Z = 1$, so führt er die Revision seines Glaubens durch

$$\begin{array}{ccc} & \textit{Bayes-Formel} & \\ p(A) & \longrightarrow & p(A|B) \\ o \downarrow & & \uparrow o^{-1} \\ \hat{p}(A) & \longrightarrow & \hat{p}(A|B) \\ & \textit{Revisionsfaktor} & \end{array}$$

Abbildung 4.2: Zusammenhang zwischen der Odds-Darstellung und der Bayes-Formel

Multiplikation des Revisionsfaktors 1.5 als $\hat{p}_{Bob}(6W|1) = 1 \cdot 1.5 = 1.5$ durch. Damit schätzt Bob die Wahrscheinlichkeit, dass Anna einen Sechserwürfel benutzt, mit $p_{Bob}(6W|1) = o^{-1}(1.5) = 40\%$ ein. Zu genau demselben Ergebnis sind wir bei der direkten Anwendung der Bayes-Formel gelangt.

Den Zusammenhang zwischen der Transformation in die Odds-Darstellung und der Revision mit der Bayes-Formel und dem Revisionsfaktor zeigt Abbildung 4.2. Bei Anwendung der Transformationsfunktion o gestaltet sich für Bob die Glaubensrevision als einfache Multiplikation mit dem Revisionsfaktor. Das Überführen einer subjektiven Wahrscheinlichkeit in die Odds-Darstellung ist also insbesondere dann von Vorteil, wenn mehrere Revisionen des Glaubens zu erwarten sind.

Ein Problem stellt für die Bayes-Formel die Berücksichtigung von Beobachtungen dar, die aufgrund des Glaubens als *unmöglich* angesehen werden. Ein Beispiel hierfür ist wie folgt: Bob ist sich sicher, dass Anna den Viererwürfel benutzt. Daher ist $p_{Bob}(6W) = 0\%$ und $\hat{p}_{Bob}(6W) = \infty$. Wie revidiert er seinen Glauben, wenn er eine Augenzahl von fünf oder sechs beobachtet? Gemäß seines Glaubens wird ein solches Ereignis als unmöglich eingestuft. Dies schlägt sich in der Bayes-Formel dadurch nieder, dass für die posteriore Wahrscheinlichkeit $\frac{0}{0}$ vorgeschlagen wird. In der Odds-Darstellung kommt der Konflikt durch die Multiplikation $\infty \cdot 0$ zum Ausdruck. Die Vorschriften der Glaubensrevision müssen also explizit vorgeben, wie mit unmöglichen Beobachtungen umgegangen wird. Zu solchen Beobachtungen kann es immer dann kommen, wenn der Revisionsfaktor null oder unendlich ist.

Theorie der verteilten Systeme. In der Literatur sind Problemstellungen bekannt, die die Eigenschaften verteilter Systemen verdeutlichen. Aus der Sicht dieser Arbeit sind hierbei das Problem byzantinischer Generäle und das Problem eines koordinierten Angriffs von Interesse.

Das *Problem byzantinischer Generäle* befasst sich damit, dass sich Einheiten in einem verteilten System zu Aussagen festlegen können, die sich untereinander widersprechen. Die Problembeschreibung ist wie folgt [LSP82]: Ein General gibt seinen beiden Lieutenants jeweils einen Befehl, der sie zum Angriff oder Rückzug auffordert. Dabei können die Lieutenants ihre Befehle nicht gegenseitig einsehen, da sie separat erteilt worden sind. Ein gewissenhafter General gibt beiden Lieutenants denselben Befehl. Um sicherzugehen, dass der General gewissenhaft ist, müssen die Lieutenants sich also untereinander von ihren Befehlen berichten. Allerdings können die Lieutenants beim Bericht über ihren Befehl lügen. Erhält zum Beispiel Lieutenant A den Befehl zum Angriff und einen Bericht von Lieutenant B, dass ihm der Rückzug befohlen wurde, so kann Lieutenant A nicht feststellen, ob der General sich widersprechende Befehle gegeben oder Lieutenant B gelogen hat. Das Problem besteht also darin, dass eine Einheit, die inkonsistente Aussagen (im Beispiel Befehle) trifft, von den anderen nicht als solche identifiziert werden kann. Dieses Problem ist lösbar für den Fall, dass Aussagen nicht-abstreitbar sind. Wenn zum Beispiel der General sein Befehl unterschreibt und die Lieutenants die unterschriebenen Befehle

austauschen, so ist ein gewissenhafter General von einem nicht gewissenhaften unterscheidbar. In verteilten Systemen ist die Nichtabstreitbarkeit demnach unabdingbar, um Einheiten, die sich fehlverhalten, identifizieren zu können.

Eine weitere Problemstellung in verteilten Systemen ist das *Problem eines koordinierten Angriffs* [HM84]. Gegeben sind die Generäle zweier Armeen, die einen Feind gemeinsam angreifen möchten. Dabei will jeder General absolut vermeiden, dass er den Feind alleine angreift. Um ihre Entscheidung zum Angriff zu koordinieren, können die Generäle Boten austauschen. Ein offensichtlicher Ansatz besteht darin, dass General A General B meldet, dass er angreifen wird. Das Problem liegt darin, dass der Bote, der diese Nachricht trägt, unter Umständen (zum Beispiel durch Feindeseinwirkung) nicht zu General B gelangen kann. Nach Absenden seiner Nachricht kann General A sich daher nicht sicher sein, dass General B seine Nachricht erhalten hat und deswegen angreifen wird. Um einen einseitigen Angriff zu vermeiden, kann General A also nicht den Feind angreifen. Dieses Problem wird auch dadurch nicht gelöst, dass General B eine Bestätigung der Nachricht versendet. In diesem Fall ist sich nämlich General B unsicher, ob General A die Bestätigung erhält und deswegen angreifen wird. Daher wird das ursprüngliche Problem durch das Versenden von Bestätigungen nur verschoben aber nicht gelöst. Unabhängig davon, wie oft sich die Generäle ihre Nachrichten bestätigen, gilt, dass zumindest derjenige General unsicher über den Angriff ist, der die letzte Nachricht sendet. Um einen Angriff unter Sicherheit auszuführen, ist aus der Sicht von General A das folgende Wissen erforderlich:

1. General B wird angreifen.
2. General B weiß, dass General A angreifen wird (sonst gilt (1) nicht).
3. General B weiß, dass General A weiß, dass General B angreifen wird (sonst gilt (2) nicht).
4. Diese Rekursion wird unendlich durchgeführt.

Sind diese Bedingungen erfüllt, so sprechen wir von einem gemeinsamen Wissen (engl.: common knowledge) der Generäle. Das Problem eines koordinierten Angriffs macht also deutlich, dass gemeinsames Wissen in einem verteilten System nicht durch das Versenden von einer endlichen Zahl von Nachrichten erreicht werden kann.

4.3 Zusammenfassung

In diesem Kapitel wurden die Grundlagen dieser Arbeit eingeführt. Es handelt sich bei ihnen vor allem um die Konzepte der Spieltheorie. Sie befasst sich mit der Frage, wie sich autonome Einheiten, die ihren eigenen Vorteil suchen, unter vorgegebenen Rahmenbedingungen verhalten. Es wurde zunächst darauf eingegangen, wie diese Rahmenbedingungen als Spiel zu modellieren sind und wie ein solches Spiel zu analysieren ist. Anschließend haben wir uns mit den Erweiterungen der Spieltheorie befasst, die eine asymmetrische Verteilung von Informationen zwischen den Spielern berücksichtigen. Aufgrund dieser Informationsasymmetrie sehen Spieler das Ausführen gewisser Handlungen als Möglichkeit an, ihren Typ anderen Spielern zu signalisieren. Im Ladenkettenspiel kommt es aus diesem Grunde dazu, dass selbst Handlungen, deren Wahl an sich irrational ist, von rationalen Spielern ausgeübt werden. Abschließend sind wir auf die evolutionäre Spieltheorie eingegangen. Ausgehend von einer Zuordnung von Strategien zu Einheiten untersucht sie, welche Strategien sich bei mehrmaliger Kooperation zwischen den Einheiten untereinander durchsetzen. Es zeigte sich, dass das Modell der Kooperation der evolutionären Spieltheorie in

mancher Hinsicht äquivalent ist mit dem Systemmodell von Informationssystemen wie dem des Campus-Szenarios.

Außer der Spieltheorie wurden einige weitere Grundlagen vorgestellt. Hierzu sind wir zunächst darauf eingegangen, welche Arten von Unsicherheit es gibt und wie die Glaubensbildung auf sie eingeht. Für die Glaubensrevision ist der probabilistisch fundierte Ansatz der Bayes-Formel eingeführt worden. Zu seinem Einsatz empfiehlt sich die Transformation subjektiver Wahrscheinlichkeiten in die Odds-Darstellung, so dass die Glaubensrevision durch eine einfache Multiplikation mit einem Revisionsfaktor durchgeführt werden kann. Abschließend sind wir auf die Probleme der byzantinischer Generäle und des koordinierten Angriffs eingegangen. Sie zeigen die Bedeutung von Nichtabstreitbarkeit und globalem Wissen für verteilte Systemen auf.

Kapitel 5

里仁為美 擇不處仁 焉得知

"Es ist die Güte, die der Nachbarschaft ihre Schönheit verleiht. Jemand, der in seiner Wahl frei ist, es aber nicht vorzieht, unter den Guten zu leben – wie kann er weise genannt werden?"
(Gespräche und Aussprüche des Konfuzius, 4.1)

Systementwurf und Autonomie

Ein entscheidendes Kennzeichen für das Informationssystem des Campus-Szenarios besteht darin, dass die Teilnehmer autonom sind. Sie entziehen sich also jeglicher äußerer Kontrolle. Die Autonomie drückt sich darin aus, dass die menschlichen Benutzer selbst entscheiden, welche Software sie auf ihren Geräten installieren. Diese Wahl bestimmt über das Verhalten der Geräte im Informationssystem. Dadurch stellt sich die Frage, inwiefern sich ein solches Informationssystem überhaupt von einem Außenstehenden entwerfen lässt. Eine Klärung dieser Frage ist für die weiteren Kapitel dieser Arbeit, in denen ein solcher Systementwurf durchführt geführt wird, unabdingbar.

Dieses Kapitel befasst sich mit dem Gegensatz zwischen Systementwurf und Autonomie und den Folgen, die daraus entstehen. Die Grundzüge dieses Gegensatzes werden in Abschnitt 5.1 dargestellt. Es folgt in Abschnitt 5.2 eine Analyse, welche Hindernisse es für Manipulation der Systemsoftware gibt. Diese Hindernisse beeinflussen maßgeblich die autonomen Teilnehmer in ihrer Wahl der verwendeten Software. Dies wird in Abschnitt 5.3 berücksichtigt, um ein Modell der am Informationssystem teilnehmenden Einheiten zu erarbeiten. Die Eigenschaften dieses Modells führen in Abschnitt 5.4 zu Schlussfolgerungen darüber, wie der Systementwurf im Hinblick auf die Autonomie der Teilnehmer gestaltet werden muss.

5.1 Einführung

Dieser Abschnitt stellt die zentralen Begriffe dieses Kapitels, die Autonomie und den Systementwurf, gegenüber. Zu diesem Zweck werden diese beiden Punkte einzeln besprochen und ihre Eigenschaften herausgestellt. Die Unterschiedlichkeit ihrer Eigenschaften führt dazu, dass ein Gegensatz zwischen der Autonomie und dem Systementwurf entsteht.

Autonomie. Jeder menschliche Benutzer entscheidet autonom darüber, ob und mit welcher Software sein Gerät am Informationssystem teilnimmt. Bei der Software kann er zwischen der

originalen Systemsoftware und manipulierten Versionen von ihr wählen. Es gibt also insgesamt drei Alternativen für die autonome Entscheidung des Benutzers:

1. *Keine Teilnahme:* Der Benutzer installiert keine Systemsoftware auf seinem Gerät. Er nimmt daher nicht am Informationssystem teil.

2. *Teilnahme mit originaler Systemsoftware:* Der Benutzer installiert die originale Systemsoftware auf sein Gerät und nimmt dadurch am Informationssystem teil.

3. *Teilnahme mit einer manipulierten Version der Systemsoftware:* Der Benutzer erstellt eine manipulierte Version oder übernimmt sie von anderen. Diese installiert er auf seinem Gerät und nimmt so am Informationssystem teil.

Der Autonomiebegriff beschränkt sich allerdings nicht auf die menschlichen Benutzer. Aus der Sicht des Informationssystems sind die teilnehmenden Einheiten, als Vertreter ihrer menschlichen Prinzipale, selbst autonom. Der Autonomiebegriff bezieht sich bei Einheiten darauf, dass sich die Einheiten des Informationssystems beliebig verhalten können. Insbesondere ist keine Einheit in der Lage, andere Einheiten zu gewissen Handlungen zu zwingen. Dies ergibt sich direkt daraus, dass die Einheiten nur von ihren jeweiligen menschlichen Prinzipalen kontrolliert werden können.

Systementwurf. Der Entwurf von Informationssystemen wie das des Campus-Szenarios unterscheidet sich grundsätzlich vom Entwurf konventioneller Informationssysteme. Dies liegt daran, dass der Initiator des Informationssystems (im Campus-Szenario die Universität) nicht selbst für den Betrieb des Systems sorgt, sondern zu diesem Zweck auf die Informationsgeräte der Benutzer zurückgreifen muss. Der Entwurf des Informationssystems besteht also lediglich im Entwurf der originalen Systemsoftware, die an interessierte Benutzer verteilt wird. Unter Systementwurf verstehen wir daher im Folgenden den Entwurf der originalen Systemsoftware.

Die originale Systemsoftware besteht aus unterschiedlichen Komponenten. Zum einen besteht die Systemsoftware aus einer funktionalen Basis. Das sind diejenigen Komponenten, die für den automatisierten Austausch von Informationen und Informationsdiensten direkt benötigt werden. Diese benutzen die Techniken der Automatisierung, wie sie in Abschnitt 2.1 besprochen wurden. Diese funktionale Basis wird um die Komponenten der verteilten Vertrauensbildung erweitert, um Robustheit gegenüber Betrugsverhalten zu erhalten. Der Entwurf der verteilten Vertrauensbildung, mit dem sich Teil II dieser Arbeit befasst, stellt also einen Teil des Systementwurfs dar. Da in dieser Arbeit die funktionale Basis als vorgegeben angenommen wird, setzen wir im Folgenden den Systementwurf mit dem Entwurf der verteilten Vertrauensbildung gleich[1].

Gegensatz zwischen Systementwurf und Autonomie. Abbildung 5.1 stellt den Systementwurf in einen Zusammenhang mit der Autonomie der Teilnehmer. Der Systementwurf besteht aus dem Entwurf der originalen Systemsoftware. Diese wird interessierten Benutzern zur Installation auf ihrem Informationsgerät angeboten. Aufgrund ihrer Autonomie können sich aber die Benutzer auch für manipulierte Versionen der Systemsoftware entscheiden. Damit bestimmen letztendlich sie darüber, wie sich ihre Einheiten im Informationssystem verhalten. Die Eigenschaften des Informationssystems ergeben sich aus diesem Verhalten der einzelnen Einheiten.

Die Abbildung macht deutlich, dass Systementwurf und Autonomie sich auf unterschiedliche Ebenen beziehen:

[1] Teile des Entwurfs der verteilten Vertrauensbildung greifen invasiv auf die funktionale Basis zu. Dies betrifft die Konzeption des Transaktionsprotokolls aus Abschnitt 6.2 und 7.2. Ein solcher invasiver Eingriff ist notwendig, weil die Verfahren aus Kapitel 7 auf ein solches Transaktionsprotokoll aufbauen.

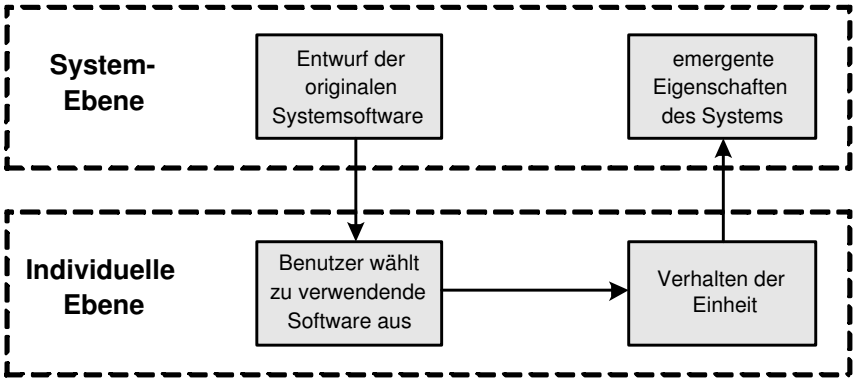

Abbildung 5.1: Zusammenspiel zwischen Systementwurf und individueller Autonomie

- *System-Ebene:* Der Systementwurf erfolgt in der Hinsicht systemweit, als die Systemsoftware allen interessierten Benutzern zur Installation angeboten wird. Auf der anderen Seite führt das Verhalten der Teilnehmer zur Emergenz der Eigenschaften des Gesamtsystems.
- *Individuelle Ebene:* Im Gegensatz zum Systementwurf ist die Autonomie eine individuelle Eigenschaft der Teilnehmer. Jeder Benutzer trifft für sich eine lokal gültige Entscheidung bezüglich der Version der verwendeten Systemsoftware. Ebenso verhält sich jede Einheit entsprechend der Vorgaben derjenigen Software, die lokal auf dem jeweiligen Gerät installiert ist.

Der Systementwurf zielt letztlich auf die Bestimmung der Systemeigenschaften. Aus der Abbildung wird aber ersichtlich, dass diese vom Systementwurf nicht direkt bestimmt werden können. Diese Entkopplung entsteht aufgrund der Autonomie der Teilnehmer. Der Systementwerfer kann nämlich nicht durchsetzen, dass seine Systemsoftware auch tatsächlich von den Geräten verwendet wird, die am Informationssystem teilnehmen. In dieser Hinsicht stehen Systementwurf und Autonomie in einem Gegensatz zueinander.

Die Rolle des Systementwerfers beschränkt sich also darauf, den Benutzern eine Systemsoftware vorzuschlagen und zur Verfügung zu stellen. Sind die Benutzer diesem Vorschlag abgeneigt (etwa weil er kein vorteilhaftes Betrugsverhalten vorsieht), wird die originale Systemsoftware nicht verwendet. In diesem Fall wird der Systementwurf nicht in die Praxis umgesetzt und gilt als gescheitert. Diese Feststellung lässt den Systementwurf zunächst als aussichtsloses Unterfangen erscheinen. Jedoch werden im folgenden Abschnitt Faktoren identifiziert, die die Benutzer tendenziell zur Verwendung der vorgeschlagenen Systemsoftware bewegen.

5.2 Hindernisse für die Manipulation der Systemsoftware

Autonomie äußert sich darin, dass die menschlichen Benutzer die Wahl haben, die originale Systemsoftware oder manipulierte Versionen davon auf ihren Geräten zu benutzen. Eine solche Wahlentscheidung ist abhängig davon, ob es für die Benutzung manipulierter Versionen Hindernisse gibt. Dieser Abschnitt geht dieser Frage nach und untersucht, wie stark etwaige Hindernisse sind.

Im Campus-Szenario erkennen wir zwei prinzipielle Möglichkeiten zur Benutzung manipulierter Versionen der Systemsoftware:

- *Erstellen einer manipulierten Version:* Ein menschlicher Benutzer (im Szenario Manuel)

ist in der Lage, die Systemsoftware in seinem Sinne zu verändern. Dadurch *erstellt* er eine manipulierte Version der Systemsoftware.

- *Übernehmen einer manipulierten Version:* Wird eine solche manipulierte Version auch anderen zum Beispiel über eine Internet-Seite zur Verfügung gestellt, so kann sie von anderen Benutzern (im Szenario Bob) *übernommen* werden.

Beide Alternativen sind jeweils mit Schwierigkeiten verbunden, die von *technischen* und *rechtlichen Hindernissen* verursacht werden. Diese werden in Abschnitt 5.2.1 und 5.2.2 vorgestellt. Abschließend werden in Abschnitt 5.2.3 die Ergebnisse der Analyse bewertet.

5.2.1 Technische Hindernisse

Im Folgenden wird auf die technischen Hindernisse für das Erstellen und Übernehmen manipulierter Versionen eingegangen. Im Gegensatz zu den Ansätzen der Hardware-basierten Manipulationssicherheit aus Abschnitt 2.2 sind diese Hindernisse nicht prohibitiv. Sie stellen aber dennoch eine nicht zu vernachlässigende Erschwernis der Manipulation dar.

Hindernisse für das Erstellen einer manipulierten Version. Folgendes Vorgehen ist erforderlich, um eine manipulierte Version zu erstellen. Nach dem Erhalt der originalen Systemsoftware wird zunächst der Binärcode der Software dekompiliert, so dass ihre Funktionsweise analysiert werden kann. Dabei wird untersucht, welche Programmstellen Verhaltensweisen bewirken, die vom menschlichen Benutzer nicht erwünscht sind. Insbesondere sind das diejenigen Programmstellen, die zu kooperativem Transaktionsverhalten führen. Aufgrund der dadurch gewonnenen Erkenntnisse wird in einem Folgeschritt das Verhalten der Software durch eine geeignete Veränderung des Programms den Bedürfnissen des Benutzers angepasst. Abschließend wird das veränderte Programm neu kompiliert, um eine manipulierte Version der Systemsoftware zu erhalten. Diesen Vorgang der Analyse, Anpassung und Synthese nennt man *Reverse Engineering*.

Es liegt auf der Hand, dass nur menschliche Benutzer mit *Expertenwissen* zum Reverse Engineering fähig sind[2]. Damit beschränkt sich der Kreis derer, die zur Erstellung einer manipulierten Version der Systemsoftware in der Lage sind, auf einen Bruchteil der am Informationssystem teilnehmenden Benutzer.

Der Aufwand des Reverse Engineering wird um mehrere Größenordnungen erhöht, wenn die Techniken der *Code-Verschleierung* [LD03] auf die originale Systemsoftware angewendet werden. Das Prinzip der Code-Verschleierung ist wie folgt: Der Binärcode wird durch eine Reihe von Transformationen verändert. Diese Transformationen sind bezüglich des Verhaltens des Programms idempotent und bewirken dadurch eine Verschleierung des Programms, ohne seine Funktionsweise zu verändern. Unter Anwendung der Code-Verschleierung gestaltet sich die Analyse der dekompilierten Systemsoftware äußerst schwierig. Damit wird der Zeitaufwand, der zum Reverse Engineering nötig ist, nachhaltig gesteigert.

Hindernisse für das Übernehmen einer manipulierten Version. Fast alle menschlichen Benutzer verfügen entweder nicht über das Expertenwissen oder nicht über die Zeit, um das Reverse Engineering durchzuführen. Allerdings gibt es eine Alternative zum eigenen Erstellen einer manipulierten Version: Ein menschlicher Benutzer, der selbst eine manipulierte Version erstellt hat, ist in der Lage, seine Version auch anderen Benutzern zur Verfügung zu stellen.

[2] Auf die Schwierigkeit des Reverse Engineering wird in der Literatur vielerorts hingewiesen [Vas02, SS02a, BH03].

Dies geschieht durch Veröffentlichung der manipulierten Version auf einer Internet-Seite oder durch eine direkte Weitergabe an den eigenen Bekanntenkreis. Ein Benutzer der dadurch Zugang zu einer manipulierten Version erhält, kann sie auf sein Informationsgerät übernehmen. In dieser Hinsicht ist nicht notwendig, dass ein Benutzer, der eine manipulierte Version der Systemsoftware benutzt, diese auch selbst erstellt hat[3].

Allerdings ist es fraglich, ob manipulierte Versionen überhaupt verfügbar gemacht werden. Derjenige, der die manipulierte Version verteilt, setzt sich nämlich einem Risiko aus, da es (wie im nächsten Abschnitt erläutert) rechtliche Hindernisse gegen sein Verhalten gibt. Es ist daher zu erwarten, dass die manipulierte Version nur auf einer schwierig auffindbaren Internet-Seite verfügbar gemacht wird. Dadurch ist das Übernehmen einer manipulierten Version mit einem gewissen *Suchaufwand* verbunden. Des Weiteren ist fraglich, ob der Ersteller einer manipulierten Version das Ergebnis seiner Arbeit auch mit anderen Benutzern zu teilen bereit ist.

Selbst wenn ein menschlicher Benutzer in der Lage ist, eine manipulierte Version zu übernehmen, existiert für ihn noch ein weiteres technisches Hindernis: Das Verhalten der übernommenen Version ist für ihn *undurchsichtig*. Insbesondere kann von einem durchschnittlichen Benutzer nicht überprüft werden, ob die übernommene Version nicht etwa die *Interessen* ihres *Erstellers* vertritt, zum Beispiel indem dessen Einheit bevorzugt behandelt wird. Ein solches Verhalten der manipulierten Version ist sogar zu erwarten, da nur dadurch ihr Ersteller einen Nutzen von ihrer Verteilung erhält. Im Extremfall kann es sich bei der übernommenen Version sogar um einen *Trojaner*[4] handeln, der das Informationsgerät des Benutzers unter seine Kontrolle bringt und für unlautere Zwecke benutzt.

Allgemeine Hindernisse für die Benutzung einer manipulierten Version. Ein weiteres technisches Hindernis erschwert sowohl das Erstellen als auch das Übernehmen einer manipulierten Version. Bei der Erstellung einer manipulierten Version wird das Verhalten der Systemsoftware abgeändert, um die Vorteile von Betrugsverhalten ausnutzen zu können. Allerdings steht dem Ersteller einer manipulierten Version kein Werkzeug zur Hand, mit dem er Effektivität des Verhaltens der manipulierten Version zu evaluieren in der Lage ist[5]. Es ist daher durchaus denkbar, dass die manipulierte Version schlechter als die originale Version abschneidet. Die *Unklarheit* darüber, ob dies zutrifft, stellt eine weitere Barriere für die Benutzung manipulierter Versionen dar.

5.2.2 Rechtliche Hindernisse

Die Benutzung einer manipulierten Version erfährt rechtliche Hindernisse in zweierlei Hinsicht. Einerseits stellt eine manipulierte Version der Systemsoftware ein *abgeleitetes Werk* dar, dessen Erstellung oder Übernahme gegen das *Urheberrecht* verstößt. Entscheidend sind hierzu im europäischen Raum die *Direktiven* der EU und speziell in Deutschland ihre Umsetzung im *Urhebergesetz*. Andererseits lässt sich in der Lizenz zur Teilnahme am Informationssystem die Erstellung und Übernahme manipulierter Versionen verbieten. Diese Lizenzen müssen von menschlichen Benutzern eingegangen werden, um die originale Systemsoftware und das Zertifikat ihrer Identität von dem Betreiber des Informationssystems zu erhalten.

[3] Auch in der Literatur wird die Beobachtung getroffen, dass ein Benutzer nicht über Expertenwissen verfügen muss, um an manipulierte Versionen zu gelangen [BH03].

[4] Ein Trojaner ist eine Software, die den Benutzer gezielt über ihre Funktionsweise falsch informiert, um von ihm installiert und ausgeführt zu werden.

[5] In der Evaluation des eigenen Ansatzes in Kapitel 9 wird genau dieser Frage nachgegangen, nämlich welche Art der Manipulation am Erfolg versprechendsten ist.

Im Folgenden wird im Detail auf die rechtlichen Hindernisse für das Erstellen und das Übernehmen manipulierter Versionen eingegangen. Der Analyse folgt eine Diskussion über die Nachhaltigkeit rechtlicher Hindernisse in selbstorganisierenden Informationssystemen mit autonomen Einheiten.

5.2.2.1 Hindernisse für das Erstellen einer manipulierten Version

In der Literatur wird darauf hingewiesen, dass Software bezüglich des Rechtsschutzes eine Sonderstellung einnimmt [Jon05]. Einerseits stellt Software ein *Werk* dar und ist damit *urheberrechtlich* geschützt. Andererseits lässt sich auch das *Patent-* und *Vertragsrecht* auf Software anwenden. Damit kann Software rechtlich auf unterschiedliche Arten geschützt werden.

Im Folgenden wird auf die Hindernisse seitens des Urheberrechts und des Vertragsrechts eingegangen. Zusätzlich wäre auch ein patentrechtlicher Schutz möglich. Er wird aber in dieser Arbeit nicht weiter betrachtet, da dazu für die originale Systemsoftware Patente angemeldet und erhalten werden müssten.

Urheberrechtliche Hindernisse. Allgemein gilt, dass Umarbeiten an einer Software nur vom *Rechtsinhaber* der Software durchgeführt werden können. Dieses Prinzip findet sich speziell in Deutschland in § 69c.2 des Urhebergesetzes (UrhG) und in der EU in § 4b der Software-Direktive [Cou91]. Im Campus-Szenario bedeutet das, dass nur die Universität selbst die originale Systemsoftware manipulieren darf. Dies liegt daran, dass eine manipulierte Version ein *abgeleitetes Werk* darstellt.

Allerdings existiert eine generelle Ausnahme, bei der Reverse Engineering dennoch rechtens ist. Als Voraussetzung dafür muss der Zweck des Reverse Engineering sein, eine Software neu zu erstellen, die mit der originalen Software *interoperabel* ist. Gemäß § 6.1 der Software-Direktive wird dazu die Analyse und Reproduktion der Schnittstellen[6] der originalen Software erlaubt. Das Reverse Engineering darf sich damit nur auf die Schnittstellen der originalen Software beziehen. In § 6.2c wird die Reproduktion von Teilen der originalen Software daher explizit untersagt. Es wird sogar gefordert, dass die neu erstellte Software in ihrem Ausdruck sich maßgeblich von der originalen Software unterscheidet. Eine Umsetzung dieses Teils der Software-Direktive findet sich in Deutschland im § 69e UrhG.

Was bedeutet das Recht auf Reverse Engineering zwecks Interoperabilität im Kontext des Campus-Szenarios? Damit die Erstellung einer manipulierte Version der Systemsoftware rechtens ist, darf *kein* Code-Teil der originalen Systemsoftware verwendet werden. Insbesondere lässt sich also nicht dadurch manipulieren, dass die originale Systemsoftware an entscheidenden Stellen angepasst wird. Damit muss nach der Analyse der originalen Systemsoftware unabhängig von ihr die manipulierte Version erstellt werden. Dies ist um ein Vielfaches aufwändiger als das gezieltes Anpassen der Systemsoftware, von dem wir bei der Besprechung der technischen Hindernisse im vorigen Abschnitt ausgegangen sind.

Auch Reverse Engineering, das auf Interoperabilität zielt, unterliegt einigen weiteren Einschränkungen, die die Möglichkeiten des Reverse Engineering weiter eingrenzen:

- Die Software, die durch das Reverse Engineering neu entstanden ist, darf nicht in *Konkurrenz* zu der originalen Software treten [Jon05, Mus98]. Das bedeutet, dass der Betrieb der neu erstellten Software die Kooperation mit der ursprünglichen Software erfordert.

[6] Bei einem verteilten Informationssystem, wie bei dem des Campus-Szenarios, fällt unter den Begriff Schnittstelle auch das Kommunikations- und Kooperationsprotokoll zwischen den Einheiten.

Insbesondere wird damit die Kooperation zwischen Instanzen der neu erstellten Software untereinander verboten. Auf das Campus-Szenario übertragen bedeutet dies, dass Geräte mit einer manipulierten Version immer nur mit Geräten, die die originale Systemsoftware benutzen, kommunizieren dürfen.

- Durch das Reverse Engineering darf die normale Benutzung der originalen Software *nicht beeinträchtigt* werden [Mus98]. Genau dies ist im Campus-Szenario aber der Fall, da die Geräte, die die originale Systemsoftware benutzen, in ihren Transaktionen betrogen werden und dadurch nicht an die gewünschten Informationen gelangen.
- In den USA gelten noch weitergehende Einschränkungen[7]. Dafür maßgeblich ist der *Digital Millennium Copyright Act* (DMCA). Ihr § 1201 legt fest, dass Reverse Engineering auch zum Zwecke der Interoperabilität nicht erlaubt ist, wenn dadurch eine technische Schutzmaßnahme überwunden werden muss [Jon05]. Ursprünglich zielt dieser Vorschrift auf Hardware-basierte Schutzmaßnahmen. Jedoch sind im weiteren Sinne auch die Techniken der Code-Verschleierung als Schutzmaßnahmen aufzufassen. Für das Campus-Szenario bedeutet dies, dass unter keinen Umständen das Erstellen einer manipulierten Version rechtens ist.

Aus den urheberrechtlichen Hindernissen gegen die Erstellung manipulierter Versionen lässt sich folgende Schlussfolgerung ziehen: Die manipulierte Version muss vollkommen neu entwickelt werden, um nicht als abgeleitetes Werk gegen geltendes Recht zu verstoßen. Außerdem dürfen diese manipulierten Versionen nicht untereinander in Kooperation treten und in Transaktionen mit der originaler Software dürfen sie nicht betrügen. Es ist offensichtlich, dass damit jegliche sinnvolle Art der Manipulation (im Campus-Szenario die Manipulationsweise Manuels) rechtswidrig ist. Weiterhin ist es fraglich, ob in Zukunft das Recht zum Reverse Engineering weiter eingeschränkt wird, so dass selbst unter Einhaltung dieser Bedingungen das Erstellen manipulierter Versionen rechtswidrig ist.

Vertragsrechtliche Hindernisse. Voraussetzung für die Erstellung einer manipulierten Version ist der Zugang zu der originalen Systemsoftware. Der Betreiber des Informationssystems (im Campus-Szenario die Universität) ist in der Lage, nur unter Auflagen diesen Zugang teilnahmeinteressierten menschlichen Benutzern zu gewähren. Dazu wird den Benutzern eine *Lizenz* vorgelegt, in die eingewilligt werden muss, um die Systemsoftware zu erhalten. Im Folgenden betrachten wir, inwiefern in der Lizenz das Erstellen einer manipulierten Version verboten werden kann.

Ist eine Lizenzvereinbarung zulässig, in der das Reverse Engineering durch den Lizenznehmer pauschal ausgeschlossen wird? Wie bei dem Urheberrecht gelten hierzu unterschiedliche Bestimmungen in den USA und der EU: Dazu besagt § 9.1 der Software-Direktive der EU, dass eine solche Lizenzvereinbarung unzulässig und, wenn dennoch abgeschlossen, nichtig ist. In den USA ist jedoch ein pauschales Verbot von Reverse Engineering in der Lizenz durchsetzbar, wie eine Untersuchung von Rechtsfällen ergibt [Jon05]. Dadurch lässt sich auch erklären, dass die Betreiber von selbstorganisierenden Informationssystemen ein Verbot von Reverse Engineering in ihre Lizenz aufnehmen[8]. Es ist nicht auszuschließen, dass auch in der EU durch zukünftige Gesetzesänderungen diesbezüglich dieselbe Rechtslage wie in den USA hergestellt wird.

[7] In der EU gibt es eine Tendenz dazu, die Rechtmäßigkeit von Reverse Engineering in Richtung des in den USA geltenden Rechts weiter einzuschränken. Dies zeigt sich besonders an der Durchsetzungsdirektive [Cou04].

[8] In der Lizenzvereinbarung von KaZaA, einem Vertreiber einer Systemsoftware für selbstorganisierende Informationssysteme, findet sich zum Beispiel folgendes [KaZ05]:

5.2.2.2 Hindernisse für das Übernehmen einer manipulierten Version

Eine manipulierte Version kann nur dann übernommen werden, wenn sie verfügbar ist. Da für das *Verbreiten* von Software im Vergleich zum Reverse Engineering weitergehende rechtliche Bestimmungen gelten, gehen wir im Folgenden zunächst auf diese ein. Anschließend werden die rechtlichen Hindernisse für einen menschlichen Benutzer untersucht, der eine manipulierte Version *übernehmen* will. Die Betrachtungen gehen davon aus, dass manipulierte Versionen rechtswidrig sind, also ihre Erstellung gegen die rechtlichen Einschränkungen des Reverse Engineering verstößt. Gemäß Abschnitt 5.2.2.1 trifft dies nämlich auf jede sinnvoll erstellte manipulierte Version zu.

Verbreitung einer manipulierten Version. Das Urheberrecht verbietet es, eine rechtswidrig erstellte manipulierte Version anderen zur Verfügung zu stellen. Entscheidend in der EU ist hierbei § 6.2 der Software-Direktive und speziell in Deutschland § 69c UrhG.

Die Situation ist beim Vertragsrecht wie folgt: In Ländern wie den USA, in denen Lizenzklauseln bezüglich des Reverse Engineering durchsetzbar sind, lässt sich die Weitergabe manipulierter Versionen in der Lizenz verbieten. Außer diesem direkten Verbot gibt es noch eine weitere Art der Lizenzvereinbarung, die auch in der EU zulässig ist. Sie besteht darin, die Systemsoftware unter der Klausel der *Vertraulichkeit* (engl.: confidentiality) weiterzugeben [Mus98, Jon05]. Eine solche Klausel ist zulässig, solange die Systemsoftware nicht offen verkauft wird. Damit greift diese Klausel auch im Campus-Szenario. Sie erzwingt, dass Informationen, die aus der vertraulich weiter gegebenen Software gewonnen worden sind, nicht an Dritte weiter gegeben werden. Im weiteren Sinne bezieht sich diese Einschränkung auch auf die manipulierten Versionen selber, die mit Hilfe dieser Informationen erstellt worden sind. Wir ziehen also den Schluss, dass die Weitergabe manipulierter Versionen (auch von Versionen, die rechtskonform erstellt worden sind) gegen das Vertragsrecht verstößt.

Übernehmen einer manipulierten Version. Ein menschlicher Benutzer übernimmt eine manipulierte Version, indem er sie von einer Internet-Seite herunterlädt oder von jemand anderem auf sein Gerät kopiert. Beide Vorgänge stellen eine *Vervielfältigung* der manipulierten Version dar. Da diese gemäß Abschnitt 5.2.2.1 als abgeleitetes Werk erstellt ist, verstößt diese Vervielfältigung gegen das Urheberrecht [Hof02]. Voraussetzung für diesen Verstoß ist allerdings, dass die manipulierte Version als *offensichtlich rechtswidrig hergestellte* Software (§ 53 UrhG) erkennbar ist. Das Wort *offensichtlich* ist allerdings dahingehend zu interpretieren, dass ein verständiger Betrachter die rechtswidrige Erstellung zu erkennen in der Lage ist [ND04]. Ob subjektiv für den einzelnen Benutzer diese Erkennbarkeit gegeben ist, ist also nicht von Belang. Es liegt also am Herausgeber der originalen Systemsoftware, die Benutzer für die Existenz von manipulierten

- § *3.2:* "Except as expressly permitted in this Licence, you agree not to reverse engineer, de-compile, disassemble, alter, duplicate, modify, rent, lease, loan, sublicense, make copies, create derivative works from, distribute or provide others with the Software in whole or part, transmit or communicate the application over a network."
- § *3.4:* "You may not use, test or otherwise utilize the Software in any manner for purposes of developing or implementing any method or application that is intended to monitor or interfere with the functioning of the Software."
- § *3.5:* "You may not through the use of any third party software application, alter or modify the values stored by the Software in your computer's memory, on your computer's hard disk, or in your computer's registry, or, with the exception of completely uninstalling the Software, otherwise modify, alter or block the functioning of the Software."

Versionen der Systemsoftware zu sensibilisieren. Für das Campus-Szenario bedeutet das, dass die Universität den Studenten vermitteln muss, dass nur direkt von ihr bezogene Systemsoftware zur Teilnahme am Informationssystem berechtigt. In einem solchen Fall ist ein verständiger Student in der Lage, manipulierte Versionen als rechtswidrig zu erkennen.

Ein vertragsrechtliches Verbot gegen die Übernahme manipulierter Versionen lässt sich hingegen nicht erreichen. Dies liegt daran, dass ein Benutzer, der eine manipulierte Version der Systemsoftware bezieht, mit dem Urheber der originalen Systemsoftware keine vertragliche Beziehung eingeht und somit keiner Lizenzvereinbarung zustimmen muss. Die Situation ändert sich allerdings, wenn ein Benutzer eine übernommene manipulierte Version auch auf seinem Gerät installieren und benutzen will. Darauf wird im folgenden Abschnitt eingegangen.

5.2.2.3 Allgemeine Hindernisse für die Benutzung einer manipulierten Version

In den vorigen Abschnitten sind die rechtlichen Hindernisse diskutiert worden, die sich speziell auf das Erstellen oder auf das Übernehmen manipulierter Versionen der Systemsoftware beziehen. Unabhängig davon, wie ein menschliche Benutzer in den Besitz einer manipulierten Version gelangt, gibt es darüber hinaus ein weiteres rechtliches Hindernis, ein manipulierte Version auf einem Informationsgerät zu installieren und damit am Informationssystem teilzunehmen. Im Folgenden wird auf dieses Hindernis eingegangen.

Der vertragsrechtliche Spielraum des Betreibers des Informationssystems (im Campus-Szenario die Universität) stellt sich wie folgt dar: **(1)** Er verteilt die originale Systemsoftware an die Benutzer, die am Informationssystem teilnehmen wollen. Es besteht daher die Möglichkeit, die Bereitstellung der originalen Systemsoftware mit einer Einwilligung des Benutzers in eine Lizenzvereinbarung zu koppeln. In den vorigen Abschnitten wurden die daraus entstehenden vertragsrechtlichen Hindernisse für die Erstellung und Verteilung manipulierter Versionen bereits aufgezeigt. **(2)** Gemäß Abschnitt 2.1 ist der Betreiber auch dafür zuständig, einem teilnahmewilligen Benutzer ein *Zertifikat* seiner *Identität* auszustellen. Die Vergabe des Zertifikats stellt damit ein Mittel dar, die Teilnahme am Informationssystem Einschränkungen zu unterwerfen. Dazu ist die Vergabe eines Zertifikats an eine Lizenzvereinbarung zu koppeln, die die Benutzung des Zertifikats in Verbindung mit manipulierten Versionen der Systemsoftware verbietet.

Für einen Benutzer existieren zwei prinzipielle Möglichkeiten, am Informationssystem teilzunehmen, ohne in die Lizenzvereinbarung des Zertifikats seiner Identität einwilligen zu müssen:

- *Verzicht auf jegliches Zertifikat:* Fehlt einem Benutzer das Zertifikat, wird sein Gerät von den anderen Einheiten des Informationssystems nicht als möglicher Kooperationspartner akzeptiert. Dies liegt insbesondere daran, dass das Fehlen des Zertifikats ein glaubhaftes Signal dafür ist, dass eine manipulierte Version der Systemsoftware verwendet wird.

- *Übernehmen des Zertifikats anderer Benutzer:* Welche Situation ergibt sich, wenn ein altruistischer Benutzer das Zertifikat seiner Identität anderen Benutzern zur Verfügung stellt? Durch die Übernahme dieses Zertifikats sind Benutzer in der Lage, auch ohne Einwilligung in die Lizenzvereinbarungen am Informationssystem teilzunehmen. Allerdings wird im Informationssystem zwischen den Einheiten dieser Benutzer nicht unterschieden, da sie sich mit derselben Identität, nämlich der des altruistischen Benutzers, ausweisen. Damit fällt Betrugsverhalten durch eine dieser Einheiten auch auf die anderen Einheiten zurück. Durch die Verfahren der verteilten Vertrauensbildung, die in den nächsten Kapiteln entworfen werden, führt die Verwendung Zertifikate anderer daher zu einem Ausschluss aus den Möglichkeiten zur Kooperation im Informationssystem.

Keine der beiden Möglichkeiten stellt einen gangbaren Weg dar. Damit muss jeder menschliche Benutzer in die Lizenzvereinbarung des Zertifikats seiner Identität einwilligen, um am Informationssystem teilnehmen zu können. Durch diese Einwilligung ist die Benutzung manipulierter Versionen der Systemsoftware rechtswidrig.

5.2.2.4 Nachhaltigkeit rechtlicher Hindernisse

Die Verwendung manipulierter Versionen der Systemsoftware verstößt mehrfach gegen das Urheber- und Vertragsrecht. Allerdings ist noch zu klären, warum ein solcher Rechtsverstoß *negative Folgen* für einen Benutzer hat. Nur dann können wir von rechtlichen Hindernissen für die Verwendung manipulierter Versionen sprechen.

Rechtswidriges Verhalten kann nicht im Informationssystem selbst bestraft werden. Dies liegt an seiner Selbstorganisation und der Autonomie der teilnehmenden Einheiten. Negative Folgen für Rechtsverstöße müssen also aus dem *Umfeld* des Informationssystems entstehen.

Die Eigenschaften des Rechtssystems unterscheiden sich *grundsätzlich* von denen des Informationssystems. Einerseits gibt es im Rechtssystem zentrale Instanzen, die Rechtsverletzungen gezielt verfolgen. Andererseits sind die menschlichen Benutzer gegenüber dem Rechtssystem nicht autonom, da sie von diesen zentralen Instanzen belangt werden können. Durch das Rechtssystem sind daher im Gegensatz zum Informationssystem negative Folgen für rechtswidriges Verhalten durchsetzbar. Entscheidend ist hierfür in der EU die Durchsetzungsdirektive [Cou04] und ihre Umsetzung in Deutschland in § 69f UrhG. Sie schreibt vor, dass **(1)** die Informationsgeräte, auf die manipulierte Versionen der Systemsoftware benutzt werden, konfisziert (§ 7) und außer Betrieb gesetzt (§ 9) werden und **(2)** der rechtsverletzende Benutzer für die Kosten des verursachten Schadens (§ 13) und die Durchsetzungs- (§ 10) und Gerichtskosten (§ 14) aufkommen muss. Dadurch wird die Nachhaltigkeit rechtlicher Hindernisse gewährleistet.

5.2.3 Bewertung

Die technischen und rechtlichen Hindernisse zur Benutzung manipulierter Versionen werden in Tabelle 5.1 zusammengefasst. Dabei wird entsprechend der vorangehenden Abschnitte zwischen Hindernissen zur Erstellung und zur Übernahme manipulierter Versionen unterschieden. Diejenigen Hindernisse, die sich sowohl auf das Erstellen als auch das Übernehmen beziehen, sind in der Tabelle gesondert ausgezeichnet.

Technische und rechtliche Hindernisse *ergänzen* sich gegenseitig. Das wird besonders für den Ersteller einer manipulierten Version deutlich. Wenn er der Aufwand des Reverse Engineering gering halten will, ändert er die originale Systemsoftware nur an den entscheidenden Stellen ab. In diesem Fall ist die resultierende manipulierte Version ein abgeleitetes Werk und verstößt gegen das Urhebergesetz. Um dieses Problem zu umgehen, müsste der Ersteller einer manipulierten Version die komplette Systemsoftware neu schreiben. Durch den Einsatz von Code-Verschleierung in der originalen Systemsoftware sind die technischen Hindernisse dafür jedoch kaum zu überwinden.

Die technischen und rechtlichen Hindernisse führen zu einem *nicht vernachlässigbaren* Aufwand für einen menschlichen Benutzer, der eine manipulierte Version der Systemsoftware benutzt. Allerdings kann für den Entwurf des Informationssystems nicht davon ausgegangen werden, dass Manipulation unmöglich ist. Eben in diesem Punkt unterscheiden sich die in diesem Abschnitt aufgeführten Hindernisse von den Techniken der Hardware-basierten Manipulationssicherheit aus Abschnitt 2.2: Der Aufwand der Manipulation alleine reicht nicht aus, um die Verwendung der originalen Systemsoftware zu bewirken. Dies erschwert den Systementwurf, da im Gegensatz zu den Systemen mit Hardware-basierter Manipulationssicherheit individuell irrationales Verhalten

Tabelle 5.1: Hindernisse zur Benutzung einer manipulierten Version der Systemsoftware

Methode der Manipulation	**Technische Hindernisse**	**Rechtliche Hindernisse**
Erstellen manipulierter Versionen	– Reverse Engineering erfordert Expertenwissen – durch Techniken der Code-Verschleierung sehr aufwändig	– Manipulierte Version verletzt als abgeleitetes Werk das Urhebergesetz – selbst Neu-Erstellung unterliegt Auflagen – in den USA: Verbot in der Lizenzvereinbarung möglich
Übernehmen manipulierter Versionen	– erfordert Verfügbarkeit manipulierter Versionen – undurchsichtig, ob manipulierte Version die Interessen des Erstellers vertritt	– Weitergabe manipulierter Version verstößt gegen Lizenz – Vervielfältigung abgeleiteter Werke verstößt gegen Urhebergesetz
Erstellen/ Übernehmen manipulierter Versionen	– unklar, ob Verhalten der manipulierten Version Erfolg haben wird	– Teilnahme am System setzt Zertifikat eigener Identität voraus – Lizenz der Zertifikatsvergabe verbietet die Benutzung manipulierter Versionen

nur bedingt durchsetzbar ist. Als Folge davon führt der Systementwurf dieser Arbeit zusätzlich Verfahren zur verteilten Vertrauensbildung ein.

5.3 Modell der Einheiten unter Autonomie

Die Autonomie der menschlichen Benutzer drückt sich darin aus, dass sie selbst entscheiden können, welche Software sie auf ihren Geräten installieren. Diese Entscheidung wird von den technischen und rechtlichen Hindernissen für die Verwendung manipulierter Versionen der Systemsoftware beeinflusst. Damit sind die menschlichen Benutzer zwar weiterhin autonom, jedoch zahlt es sich für sie unter Umständen nicht aus, durch die Verwendung einer manipulierten Version das Verhalten ihrer Geräte gezielt zu beeinflussen.

Dieser Abschnitt befasst sich mit den Auswirkungen dieser Überlegungen auf das Modell der Einheiten, die am Informationssystem teilnehmen. Abschnitt 5.3.1 unterscheidet drei verschiedene Typen von Einheiten, unter denen der menschliche Prinzipal aufgrund seiner Autonomie wählen kann. Außerdem werden die Faktoren vorgestellt, die eine solche Wahl maßgeblich bestimmen. Von der Typwahl, die jeder Prinzipal individuell für seine Einheit trifft, werden die Eigenschaften

des Gesamtsystems bestimmt. Abschnitt 5.3.2 befasst sich mit dem emergenten Verhalten dieses Gesamtsystems. Anschließend wird in Abschnitt 5.3.3 auf die Gestaltungsmöglichkeiten manipulierter Versionen der Systemsoftware eingegangen. Die Untersuchung ermöglicht Aussagen über die Eigenschaften von einem der drei Typen von Einheiten. Das sich ergebende Modell der Einheiten wird in Abschnitt 5.3.4 mit dem evolutionären Modell von Einheiten verglichen. Trotz einiger Gemeinsamkeiten bestehen einige substantielle Unterschiede zu diesem Modell.

5.3.1 Typisierung der Einheiten und Typwahl des Prinzipals

Dieser Abschnitt bildet die Grundlage für das Modell der Einheiten. Ausgehend vom Entscheidungsraum des autonomen Benutzers wird eine Typisierung der Einheiten vorgestellt. Anschließend werden die Faktoren vorgestellt, die eine Wahl zwischen diesen Typen bestimmen. Diese werden untereinander in Verbindung gebracht, um zu antizipieren, wie ein rationaler Benutzer den Typ seiner Einheit wählt.

Typisierung der Einheiten. Für einen menschlichen Benutzer gibt es nach Abschnitt 5.1 insgesamt drei Alternativen bezüglich seiner Teilnahme am Informationssystem. Sie bestehen aus seiner Wahl darüber, ob und mit welcher Software sein Gerät am Informationssystem teilnimmt. Seine autonome Entscheidung hierüber bestimmt den *Typ* seiner Einheit. Wir unterscheiden also drei Typen:

1. *Normative Einheiten:* Der Benutzer verwendet die originale Systemsoftware auf seinem Gerät, um am Informationssystem teilzunehmen.

2. *Strategische Einheiten:* Der Benutzer erstellt eine manipulierte Version oder übernimmt sie von anderen. Diese verwendet er auf seinem Gerät, um am Informationssystem teilzunehmen.

3. *Nicht-teilnehmende Einheiten:* Der Benutzer installiert keine Systemsoftware auf seinem Gerät. Er nimmt daher nicht am Informationssystem teil.

Eine Sonderstellung nimmt der dritte Typ ein. Bei ihm handelt es sich um Einheiten, die durch die Weigerung des Benutzers, am Informationssystem teilzunehmen, dem System entzogen werden. Die Berücksichtigung dieses Typs von Einheiten ist wichtig, da die Zahl solcher nicht-teilnehmenden Einheiten das Potential anzeigt, das dem Informationssystem aufgrund nicht wünschenswerter Systemeigenschaften abhanden gekommen ist.

Die Bezeichnungen der zwei weiteren Typen, normativ und strategisch, kennzeichnen das Verhalten, das für die Einheiten dieser Typen jeweils charakteristisch ist. Unter Verwendung der originalen Systemsoftware verhält sich eine Einheit, wie es vom Systementwurf vorgesehen ist. Dadurch hält sie sich an die Normen des Systems, wie sie in Abschnitt 5.4.2 vorgestellt werden. In dieser Hinsicht ist eine solche Einheit *normativ*. Wird hingegen eine manipulierte Version der Systemsoftware verwendet, so ist es möglich, dass das Verhalten der Einheit von diesen Normen abweicht. Eben hierin liegt die Motivation für die Manipulation der originalen Systemsoftware. Der Ersteller einer manipulierten Version versucht gezielt, Schwachstellen der originalen Systemsoftware zu seinem Vorteil auszunutzen. Dazu legt er bei der Erstellung einer manipulierten Version eine Strategie fest, unter welchen Umständen ihr Verhalten von dem der originalen Systemsoftware abweicht. In dieser Hinsicht sind die Einheiten, bei denen eine manipulierte Version der Systemsoftware verwendet wird, *strategisch*.

Typwahl durch den menschlichen Prinzipal. Ein menschlicher Benutzer wählt den Typ seiner Einheit danach aus, welche Vor- und Nachteile die einzelnen Typen ihm jeweils bieten. Im Folgenden werden die Faktoren herausgearbeitet, die entscheidend für die Typwahl sind.

Ein Benutzer nimmt nur dann am Informationssystem teil, wenn es für ihn von Nutzen ist. Zu diesem Zweck ist der Nutzen, der sich aus dem Zugang zu Informationen und Informationsdiensten ergibt, mit dem Aufwand zu vergleichen, den eine Teilnahme mit sich bringt. Die Differenz zwischen dem Nutzen und Aufwand ist der residuale *Individualnutzen* des Benutzers. Nur wenn dieser positiv ist, entscheidet sich der Benutzer dazu, mit seinem Gerät am Informationssystem teilzunehmen. Erscheint dem Benutzer sein Individualnutzen als negativ, so weigert er sich am Informationssystem teilzunehmen. Zum Beispiel entscheiden sich im Campus-Szenario Anna und Claude zum Austritt aus dem System, weil ihre Einheiten durch das Betrugsverhalten von Manuels und Bobs Einheit kaum mehr Nutzen aus der Kooperation ziehen können, während die Kooperation selbst aufwändig bleibt.

Entscheidet sich ein Benutzer zur Teilnahme am Informationssystem, so verbleibt ihm die Wahl der Version der verwendeten Systemsoftware. Hierfür sind die Vor- und Nachteile von normativen Einheiten gegenüber strategischen Einheiten abzuwägen.

Normative Einheiten verhalten sich immer so, wie es im Systementwurf vorgesehen ist. Im Gegensatz dazu ist es den strategischen Einheiten möglich, von diesem Verhalten abzuweichen und dadurch Gelegenheiten zu vorteilhaftem Betrugsverhalten wahrzunehmen. Die Einhaltung der Normen des Systementwurfs verursacht den normativen Einheiten daher Opportunitätskosten, die wir im Folgenden *Normativitätskosten* nennen. Sie geben die Einbußen des Individualnutzens eines Benutzers an, der auf die Verwendung manipulierter Versionen zugunsten der originalen Systemsoftware verzichtet. Eine Einschätzung der Normativitätskosten ist für den einzelnen Benutzer zwar schwierig aber nicht unmöglich. Außer der probeweisen Verwendung einer manipulierten Version stehen ihm hierfür unter Umständen auch die Erfahrungen anderer Benutzer zur Verfügung.

Für einen Benutzer bringt es jedoch nicht nur Vorteile mit sich, wenn seine Einheit strategisch und nicht normativ ist. Hierfür ist nämlich die Verwendung einer manipulierten Version der Systemsoftware notwendig, für die gemäß Abschnitt 5.2 technische und rechtliche Hindernisse bestehen. Die Kosten, die durch diese Hindernisse entstehen, nennen wir im Folgenden *Manipulationskosten*. Diese sind für verschiedene menschliche Benutzer nicht notwendigerweise gleich hoch, da die Hindernisse von ihnen unterschiedlich bewertet werden:

- *Technische Hindernisse:* Nur technisch versierte Benutzer sind in der Lage, eine manipulierte Version zu erstellen oder bei Übernahme einer manipulierten Version ihr Verhalten zu überprüfen. Die technischen Hindernisse zur Manipulation sind für solche technisch versierten Benutzer demnach weitaus kleiner als für den durchschnittlichen Benutzer. Letztere unterscheiden sich wiederum darin, wie einfach sie Zugang zu einer manipulierten Version erhalten, deren Ersteller sie als vertrauenswürdig erachten. Eine weitere Differenzierung der Benutzer ergibt sich aus deren persönlichen Vorlieben. Für manche Benutzer stellt die Erstellung einer manipulierten Version eine Herausforderung dar, deren Bewältigung ihnen Spaß bereit. Diese so genannten Hacker verbinden mit dem Überwinden der technischen Hindernisse kaum oder keine Kosten.

- *Rechtliche Hindernisse:* Auch bei den rechtlichen Hindernissen ist eine unterschiedliche Bewertung durch die menschlichen Benutzer zu erwarten. Die Einen werden von der Rechtswidrigkeit ihres Verhaltens kaum abgeschreckt, während die Anderen alleine aus moralischen Gesichtspunkten jeden Rechtsverstoß vermeiden.

Die Betrachtung zeigt also, dass die Manipulationskosten eine individuelle Eigenschaft sind, die bei den menschlichen Benutzern unterschiedlich ausgeprägt ist. Ein Benutzer entscheidet sich zu einer normativen Einheit, wenn seine Manipulationskosten die Normativitätskosten überschreiten. Konkret bedeutet dieser Fall, dass der Benutzer die Nachteile des Systementwurf-konformen Verhaltens der originalen Systemsoftware akzeptiert, um nicht die Hindernisse für die Verwendung einer manipulierten Version bewältigen zu müssen. Sind die Manipulationskosten eines Benutzers andererseits niedriger als die Normativitätskosten, so entscheidet er sich zu einer strategischen Einheit. Die heterogene Ausprägung der Manipulationskosten bewirkt hierbei, dass die Benutzer sich bei der Typwahl ihrer Einheit unterschiedlich entscheiden können.

5.3.2 Emergentes Systemverhalten

Die Typwahl eines jeden menschlichen Benutzers ist ein lokaler Vorgang, der nur für seine jeweilige Einheit Gültigkeit besitzt. Der Typ einer Einheit bestimmt auf maßgebliche Weise ihr Verhalten, das wiederum Konsequenzen für die Eigenschaften des Gesamtsystems besitzt. Es stellt sich daher die Frage, wie die Typwahl der Benutzer sich auf das Informationssystem auswirkt und mit seinen Eigenschaften zusammenhängt.

Dieser Abschnitt beschäftigt sich mit der Klärung dieser Frage. Dazu wird zunächst gezeigt, dass eine Rückkopplung besteht zwischen der Typwahl der Benutzer und den Eigenschaften des Systems. Anschließend wird die dadurch hervorgerufene Dynamik der Systemeigenschaften am Beispiel des Campus-Szenarios untersucht. Dies ermöglicht eine Analyse davon, welche Systemeigenschaften letztendlich zu erwarten sind.

Rückkopplung zwischen Typwahl und Systemeigenschaften. Jedem Benutzer steht die Wahl zwischen den drei Typen von Einheiten, nämlich normativ, strategisch und nicht-teilnehmend, offen. Aus der Wahl der einzelnen Benutzer leitet sich der Zustand des Informationssystems ab: Sei a und b der Anteil der Benutzer, die sich zu einer normativen beziehungsweise strategischen Einheit entscheiden. Damit beträgt der Anteil nicht-teilnehmender Benutzer $(1 - a - b)$. Der Systemzustand bezüglich des Typs der Einheiten ist also durch das Tupel (a, b) darstellbar. Im Folgenden wird dieser Tupel auch *Populationsstruktur* des Informationssystems genannt, da er die Anteile der einzelnen Typen von Einheiten festhält.

Die Populationsstruktur bestimmt auf maßgebliche Weise, welches Verhalten im Informationssystem vorherrscht. Ist etwa der Anteil strategischer Einheiten hoch, so kann es vermehrt zu Betrugsverhalten kommen. Dies hätte wiederum eine geringe Intensität der Kooperation zur Folge, da den Einheiten potentielle Transaktionspartner wenig vertrauenswürdig erscheinen. Es zeigt sich also, dass sich die Eigenschaften des Gesamtsystems auf emergente Weise aus der Populationsstruktur ergeben.

Die Systemeigenschaften beeinflussen wiederum die Faktoren, die die Benutzer bei ihrer Typwahl berücksichtigen. Dies gilt nicht nur für den Individualnutzen, den der Benutzer durch die Teilnahme am Informationssystem erzielt. Auch die Normativitätskosten können von den Systemeigenschaften beeinflusst werden. Dieser Zusammenhang wird unter anderem im vorigen Beispiel eines Systems mit einem hohen Anteil strategischer Einheiten deutlich[9]: Aufgrund des Vertrauensmangels kann es dort zu weniger Kooperation zwischen den Einheiten kommen. Folglich erzielen die Benutzer tendenziell einen geringeren Individualnutzen. Außerdem verringern sich mit den

[9] Diese Besprechung der Dynamik des Individualnutzens und der Normativitätskosten greift auf Teil III dieser Arbeit vor, der die hier vorgebrachten Überlegungen quantitativ belegt. Die hier vorgebrachte Überlegung dient lediglich dazu, plausibel zu machen, dass die Faktoren der Typwahl von den Systemeigenschaften abhängig sind.

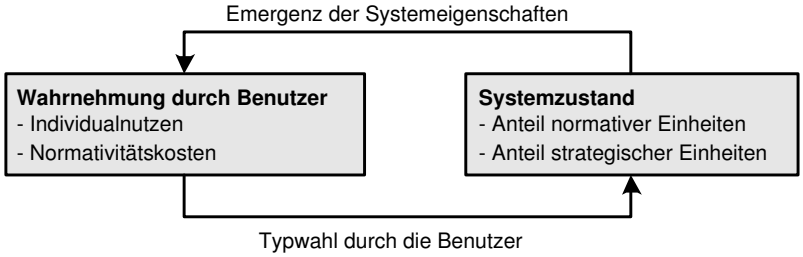

Abbildung 5.2: Rückkopplung zwischen Typwahl und Systemeigenschaften

Gelegenheiten zur Kooperation auch die zu vorteilhaftem Betrugsverhalten. Daher verkleinert sich der Vorteil der strategischen Einheiten gegenüber den normativen Einheiten. Dies schlägt sich in geringeren Normativitätskosten nieder.

Es zeigt sich also, dass es eine *Rückkopplung* zwischen der Typwahl der Benutzer und den Systemeigenschaften gibt. Diese wird in Abbildung 5.2 dargestellt. Die Entscheidung der Benutzer zum Typ ihrer Einheit fällt durch diese Rückkopplung auf sie selbst zurück. Als Ergebnis davon revidieren einige Benutzer unter Umständen ihre Typwahl. Die Populationsstruktur des Informationssystems kann sich daher dynamisch entwickeln.

Dynamik des Systems am Beispiel der Szenariobeschreibung. Die Rückkopplung zwischen Typwahl und Systemeigenschaften führt zur Dynamik der Populationsstruktur des Informationssystems. Im Folgenden untersuchen wir, wie sich diese Dynamik im Campus-Szenario niederschlägt. Dabei gehen wir wie in der Szenariobeschreibung aus Abschnitt 1.2.1 davon aus, dass im Informationssystem keine verteilte Vertrauensbildung durchgeführt wird.

Abbildung 5.3 zeigt die Dynamik der Populationsstruktur. Dazu sind die beiden Komponenten des Tupels (a, b) auf die Achsen eingezeichnet. Der grau hinterlegte Bereich stellt die Menge aller möglichen Populationsstrukturen dar. Die entscheidenden Stationen entlang der Evolution des Informationssystems sind mit den Punkten **A** bis **F** gekennzeichnet. Diese gestalten sich wie folgt:

- **A:** Alle Benutzer verwenden die originale Systemsoftware. Dies entspricht der Situation, die bei der Beschreibung des typischen Ablaufs des Szenarios vorherrscht.

- **B:** Einige Benutzer (im Szenario Manuel) erstellen eine manipulierte Version der Systemsoftware und verwenden sie. Am Informationssystem nimmt also auch eine geringe Zahl strategischer Einheiten teil.

- **C:** Eine manipulierte Version wird auch technisch weniger visierten Benutzern verfügbar gemacht. Sie wird von solchen Benutzern übernommen, deren Manipulationskosten niedrig genug sind. In der Szenariobeschreibung ist dies bei Bob der Fall. Als Folge davon steigt der Anteil der strategischen Einheiten auf Kosten der normativen Einheiten weiter an. Dieser Anstieg ist stärker als bei **B**, da die meisten Benutzer nur in der Lage sind, manipulierte Versionen zu übernehmen aber nicht selber zu erstellen.

- **D:** Durch die steigende Zahl strategischer Einheiten kommt es vermehrt zu Betrugsverhalten. Dies führt dazu, dass der Individualnutzen einiger Prinzipale normativer Einheiten

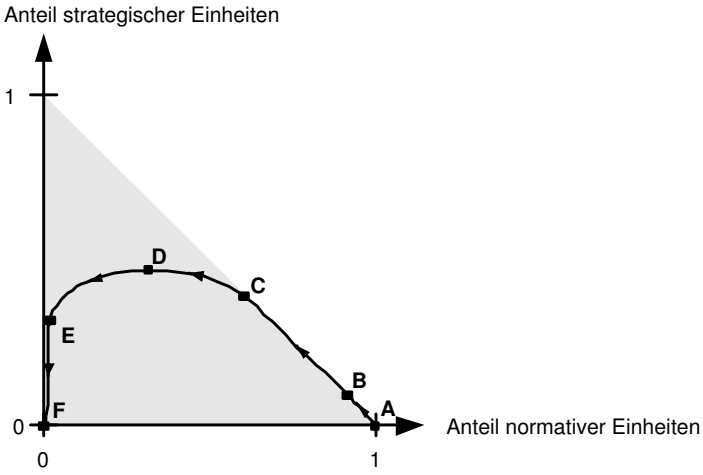

Abbildung 5.3: Degeneration des Informationssystems gemäß der Beschreibung des Campus-Szenarios

negativ wird. Als Folge davon treten Benutzer wie Anna und Claude aus dem Informationssystem aus. Ihre Manipulationskosten sind zu hoch, als dass die Verwendung einer manipulierten Version der Systemsoftware für sie eine Alternative darstellt.

- **E:** Diese Tendenz verstärkt sich, bis am Informationssystem keine normativen Einheiten mehr teilnehmen. Damit ist die adverse Selektion der Einheiten abgeschlossen. Unter diesen Umständen herrscht Betrugsverhalten im Informationssystem vor. Damit bringt auch für die verbleibenden strategischen Einheiten die Teilnahme am Informationssystem keinen Nutzen mehr.

- **F:** Letztlich treten daher auch die verbleibenden Prinzipale strategischer Einheiten aus dem Informationssystem aus.

Die Evolution des Information stellt sich als ständige Degeneration dar: Ausgehend von der Populationsstruktur $(1, 0)$, bei der sich alle Einheiten an den Systementwurf halten, entwickelt sich zwischenzeitlich ein erheblicher Anteil strategischer Einheiten. Unter den veränderten Systemeigenschaften treten zuerst die normativen und später die strategischen Einheiten aus, so dass am Ende das Informationssystem in der Populationsstruktur von $(0, 0)$ nicht mehr existiert. Wegen dieser Degeneration lag die Szenariobeschreibung in Abschnitt 1.2.1 den Schluss nahe, dass das Informationssystem nicht existenzfähig ist.

Attraktoren des Systems. Die Degeneration des Informationssystems gemäß der Szenariobeschreibung hängt eng damit zusammen, dass Betrug nicht effektiv eingedämmt wird. Mit der Einbeziehung der verteilten Vertrauensbildung in den Systementwurf ändert sich jedoch die Situation. Die Nichtexistenz des Informationssystems ist nicht mehr zwingendermaßen der einzige Attraktor des Systems. Im Folgenden untersuchen wir, unter welchen Umständen weitere Attraktoren existieren. Dies gibt Aufschluss darüber, welche Populationsstruktur in einem existenzfähigen Informationssystem vorgefunden werden kann.

Die Typwahl der Benutzer hängt maßgeblich von ihren individuellen Manipulationskosten ab. Um eine Aussage über die Attraktoren des Systems machen zu können, muss die Verteilung

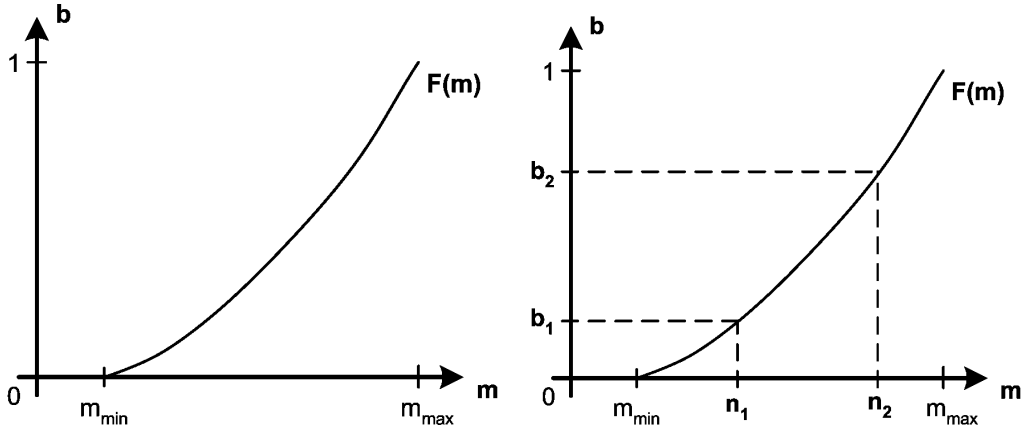

Abbildung 5.4: **(a)** Verteilung der Manipulationskosten zwischen den Benutzern; **(b)** Gemischte Gleichgewichte der Populationsstruktur

der Manipulationskosten bekannt sein: Seien m_b die Manipulationskosten des Benutzers b. Sie sind gemäß Abschnitt 5.3.1 individuell unterschiedlich. Es lässt sich daher aus den einzelnen m_b eine Verteilung F der Manipulationskosten über die Menge der Benutzer angeben. Dabei gibt $F(m)$ an, wie hoch der Anteil der Benutzer ist, deren Manipulationskosten maximal m betragen. Abbildung 5.4(a) zeigt beispielhaft einen Verlauf dieser Funktion. In der Abbildung sind die größten beziehungsweise kleinsten vorfindbaren Manipulationskosten mit $m_{max} = \max(m_b)$ und $m_{min} = \min(m_b)$ dargestellt.

Die Verteilung der Manipulationskosten gibt Aufschluss darüber, unter welchen Bedingungen sich wie viele Benutzer zur Verwendung manipulierter Versionen der Systemsoftware entscheiden. Dazu ist gemäß der Faktoren der Typwahl ein Vergleich mit der Höhe der Normativitätskosten notwendig, die mit n bezeichnet wird. Ein teilnehmender Benutzer b entscheidet sich zu einer normativen Einheit, wenn $n < m_b$ gilt. Ansonsten entscheidet er sich zu einer strategischen Einheit. Als *Stabilitätsbedingung* für das Informationssystem erhalten wir demnach die beiden folgenden Kriterien:

1. Die Prinzipale der teilnehmenden Einheiten wollen nicht aus dem System austreten. Umgekehrt bleiben auch die Prinzipale nicht-teilnehmender Einheiten dabei, nicht am System teilzunehmen. Dieses Kriterium lässt sich anhand des Individualnutzens der jeweiligen Benutzer überprüfen.

2. Die Prinzipale der teilnehmenden Einheiten bleiben bei ihrer Wahl bezüglich der verwendeten Software. Für Benutzer normativer und strategischer Einheiten muss also $n < m_b$ beziehungsweise $n > m_b$ gelten.

Wird diese Stabilitätsbedingung erfüllt, so liegt die Populationsstruktur in einem *Gleichgewicht*. Die Benutzer verändern ihre Typwahl nicht, da die Systemeigenschaften ihnen dazu keinen Anlass bieten. Auch eine gemischte Populationsstruktur aus normativen und strategischen Einheiten kann im Gleichgewicht stehen. Dies ist dann der Fall, wenn n im Bereich zwischen m_{min} und m_{max} liegt. Abbildung 5.4(b) verdeutlicht dies für zwei verschiedene Fälle, in denen die Normativitätskosten n_1 und n_2 betragen. Das Schaubild weist auf eine besondere Interpretation von $F(n)$ hin. Dieser Wert gibt den Anteil b der strategischen Einheiten an, der bei den Normativitätskosten n unter den teilnehmenden Einheiten zu erwarten ist. Die Abbildung visua-

lisiert daher den plausiblen Zusammenhang, dass eine Erhöhung der Normativitätskosten zu einer größeren Zahl von strategischen Einheiten führt. Umgekehrt zeigt dies, dass der Systementwurf für minimale Normativitätskosten sorgen muss, damit so viele Einheiten wie möglich normativ am Informationssystem teilnehmen.

Es bleibt zu untersuchen, inwiefern eine im Gleichgewicht befindliche Populationsstruktur einen Attraktor des Systemzustands darstellt. Dazu muss die Dynamik der Populationsstruktur ausgehend vom initialen Systemzustand betrachtet werden. Der Übersichtlichkeit wegen zeigen wir eine solche dynamische Betrachtung unter der Annahme, dass alle Benutzer am Informationssystem teilnehmen und sie ihre Typwahl zu gleichen Zeitpunkten revidieren. Abbildung 5.5 stellt hierfür beispielhaft die Dynamik der Populationsstruktur dar. Sie ist maßgeblich davon geprägt, dass die Normativitätskosten von den Systemeigenschaften abhängen, die sich wiederum aus der Populationsstruktur ergeben. Daher ist zusätzlich zur Verteilung der Manipulationskosten im Schaubild auch eine Funktion $N(b)$ eingezeichnet, die für eine gegebene Populationsstruktur, angegeben durch den Anteil strategischer Einheiten, die sich ergebenden Normativitätskosten zurückgibt. In der Abbildung wird beispielhaft der Fall dargestellt, dass die Normativitätskosten bei einer ausgeglichenen Mischung zwischen normativen und strategischen Einheiten am größten sind[10].

Aufgrund der Rückkopplung zwischen Normativitätskosten und Populationsstruktur ergibt sich ein iterativer Prozess, dessen Schritte sich im Schaubild folgendermaßen darstellen:

- Initial sind alle Einheiten normativ, da noch keine manipulierte Version verfügbar ist. Damit ist der initiale Anteil b_0 der strategischen Einheiten gleich Null ($b_0 = 0$). Die Normativitätskosten, die sich bei dieser rein normativen Population ergeben, sind mit $n_0 = N(b_0)$ eingezeichnet.

- Die Benutzer mit geringen Manipulationskosten ändern ihre Typwahl. Dies ist für solche Benutzer b der Fall, bei denen $m_b < n$ gilt. Damit beträgt als Folge der Anteil b_1 strategischer Einheiten $F(n_0)$. Der Übergang zwischen b_0 und b_1 wird im Schaubild durch einen vertikalen Pfeil dargestellt. Der erhöhte Anteil an strategischen Einheiten hat wiederum Auswirkungen auf die Normativitätskosten, die sich auf $n_1 = N(b_1)$ erhöhen. Dies ist im Schaubild mit einem entsprechenden horizontalen Pfeil dargestellt.

- Wird dieser Prozess fortgesetzt, so konvergiert er im Schaubild zum Punkt (n^*, b^*), für den als Fixpunkt $b^* = F(N(b^*))$ beziehungsweise $n^* = N(F(n^*))$ gilt.

In der Abbildung konvergiert der iterativen Prozess zur Populationsstruktur b^*, die die Stabilitätsbedingung erfüllt. Allerdings kann in diesem Modell der Dynamik der Populationsstruktur die Konvergenz des iterativen Prozesses nicht allgemein gewährleistet werden. Bei einem anderen Verlauf der Funktion N könnten n und b ebenso gut um einen Fixpunkt pendeln, ohne sich ihm zu nähern. Dieses Phänomen tritt jedoch nicht auf, wenn strategische Einheiten nicht wieder normativ werden können. In diesem Fall steigt der Anteil b der strategischen Einheiten monoton an, so dass wie im Beispiel die Konvergenz gesichert ist. Inwiefern ist diese Annahme gerechtfertigt? Um eine manipulierte Version der Systemsoftware zu verwenden, muss ein Benutzer b die technischen und rechtlichen Hindernisse dazu überwinden. Eben daher rühren seine Manipulationskosten m_b. Sind jedoch diese Hindernisse bereits überwunden (zum Beispiel indem eine manipulierte Version selbst erstellt wurde), so sind diese Kosten bereits aufgetreten und nicht weiter entscheidungsrelevant. Daher wird der Benutzer einer manipulierten Version sich auch bei $n < m_b$ nicht zur

[10] Einen ähnlichen Verlauf der Normativitätskosten werden wir als Ergebnis der simulativen Evaluation in der Abbildung 10.4(b) erhalten.

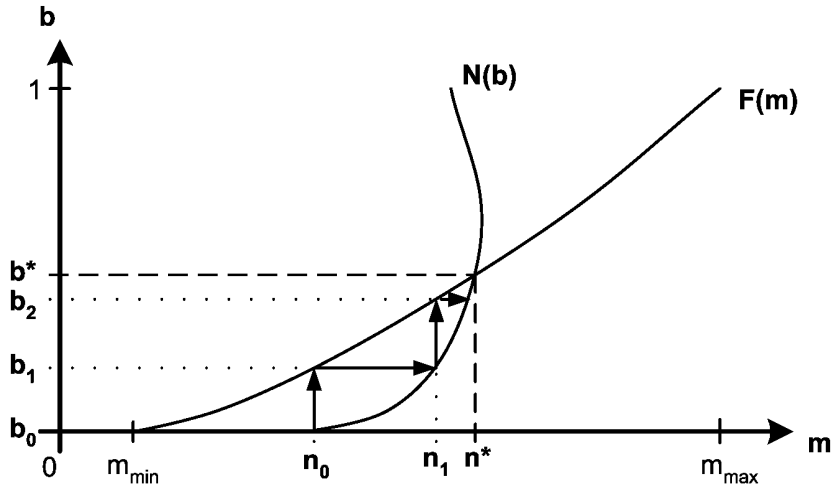

Abbildung 5.5: Dynamik und Konvergenz der Populationsstruktur

Verwendung der originalen Systemsoftware entschließen. Daraus wird ersichtlich, dass strategische Einheiten auch bei veränderten Normativitätskosten schwerlich wieder normativ werden.

Wir fassen also zusammen, dass eine der Populationsstrukturen, die die Stabilitätsbedingung erfüllen, ein Attraktor des Systems ist. Die Eigenschaften des Attraktors hängen maßgeblich von den Manipulationskosten der Benutzer und den Normativitätskosten ab. Je höher die Manipulationskosten und je niedriger die Normativitätskosten desto größer ist der Anteil der normativen Einheiten.

5.3.3 Modell strategischer Einheiten

Das Verhalten von normativen Einheiten wird von den Eigenheiten der originalen Systemsoftware bestimmt. Damit ergibt sich das Modell der normativen Einheiten direkt aus dem Systementwurf. Strategische Einheiten verhalten sich hingegen gemäß ihrer jeweilig verwendeten manipulierten Version der Systemsoftware. Um ein Modell strategischer Einheiten zu erhalten, ist also Kenntnis darüber nötig, welche Arten von manipulierten Versionen zu erwarten sind.

Dieser Abschnitt betrachtet daher die Prinzipien, die der Erstellung manipulierter Versionen der Systemsoftware zugrunde liegen. Das Vorgehen eines Erstellers solcher Versionen kann anhand der zwei folgenden Kriterien charakterisiert werden:

- *Art der Bereitstellung:* Erstellt ein Benutzer eine manipulierte Version, so kann er wählen, ob er sie anderen Benutzer zur Verfügung stellt. Wenn der Benutzer nicht nur eine manipulierte Version erstellt hat, steht für ihn zusätzlich die Entscheidung an, welche seiner Versionen er bereitstellt. Wir unterscheiden also insgesamt drei Fälle:

  - Der Ersteller verwendet seine manipulierte Version als Einziger, da er sie nicht bereitstellt.
  - Er stellt genau diejenige Version bereit, die er selbst verwendet.
  - Er hat zwei verschiedene manipulierte Versionen erstellt. Die eine verwendet er selbst und die andere stellt er anderen Benutzern bereit.

- *Rationalität der bereitgestellten Version:* Manipulierte Versionen unterscheiden sich darin, ob sie Verhalten vorschreiben, das für einen Benutzer, der sie verwendet, vorteilhaft ist oder nicht. Ist das Verhalten der Version im Sinne des Benutzers, so ist sie *individuell rational*. Wenn die Version hingegen Interessen anderer, zum Beispiel des Erstellers, verfolgt, dann ist sie *individuell irrational*. Ein weiteres Kriterium für das Vorgehen eines Erstellers einer manipulierten Version ist also, ob er anderen Benutzern eine individuell rationale oder irrationale Version bereitstellt.

Der einfachste Fall besteht darin, dass der Ersteller einer manipulierten Version sie als Einziger verwendet. Da der Ersteller dabei seine eigenen Interessen verfolgt, ist seine Version individuell rational. Ein ähnliches Bild ergibt sich, wenn ein Benutzer eine individuell rationale Version erstellt und sie Anderen zur Verfügung stellt. In diesem Fall erscheint den Benutzern die Übernahme der manipulierten Version besonders vorteilhaft, da diese Version ihre eigenen Interessen vertritt. Für beide Fälle erhalten wir also, dass strategische Einheiten sich individuell rational verhalten.

Die individuelle Rationalität von strategischen Einheiten hat jedoch zur Folge, dass sich strategische Einheiten untereinander betrügen können, auch wenn ihre jeweiligen Prinzipale dieselbe manipulierte Version der Systemsoftware verwenden. Auf den ersten Blick liegt es daher für den Ersteller einer manipulierten Version nahe, anderen Benutzern keine individuell rationalen Versionen bereitzustellen. Stattdessen sorgt er dafür, dass die Einheiten der übernahmewilligen Benutzer sich gegenseitig unterstützen und somit ein Komplott (engl.: collusion) bilden. Es bieten sich zwei prinzipielle Möglichkeiten an, wie der Ersteller manipulierter Versionen hierfür verfahren kann. Bei der nachfolgenden Besprechung dieser Möglichkeiten gehen wir davon aus, dass die übernahmewilligen Benutzer durch entsprechendes Austesten in der Lage sind, die Vorteilhaftigkeit der einzelnen manipulierten Versionen einschätzen zu können:

- *Gegenseitigkeit des Komplotts:* Als Voraussetzung dafür, dass sich die Einheiten eines Komplotts gegenseitig unterstützen, müssen sich die Teilnehmer des Komplotts als solche erkennbar machen können. Zu diesem Zweck ist in der manipulierten Version ein Mechanismus einzubauen, mit dem sich Einheiten als Teilnehmer des Komplotts ausweisen können[11]. Dieser Mechanismus wird allerdings durch weitere manipulationswillige und -fähige Benutzer wie folgt ausgenutzt: Zunächst beschaffen sie sich die manipulierte Version V, die das Bilden des gegenseitigen Komplotts veranlasst. Diese Version manipulieren sie ihrerseits. Die dabei entstehende Version V' ist derart erstellt, dass bei ihrer Verwendung nur die Vorteile aber nicht die Nachteile der Komplott-Zugehörigkeit entstehen. Konkret sieht dies so aus, dass eine Einheit mit V' sich als Teilnehmer des Komplotts ausgibt und damit von Einheiten mit V unterstützt wird, ohne allerdings diese Unterstützung zu erwidern. Aufgrund des Vorteils, den sich Einheiten durch die Verwendung von V' verschaffen, kommt es daher zur adversen Selektion der Teilnehmer des Komplotts zugunsten der Einheiten mit V'. Schlussendlich degeneriert das Komplott und ist damit nicht existenzfähig. Als Folgerung erhalten wir, dass es nicht möglich ist, ein Komplott aufzubauen, in dem sich die Einheiten gegenseitig unterstützen.

[11] Ein weitergehender Mechanismus bestünde darin, dass der Urheber eines Komplotts Normen entwirft, deren Einhaltung von den Mitgliedern des Komplotts untereinander kontrolliert wird. Dabei treten allerdings eine Reihe von Problemen für das Komplott auf, die erst im Verlauf dieser Arbeit durch die Darstellung des eigenen Systementwurfs deutlich werden. Die Besprechung eines solchen Norm-setzenden Komplotts erfolgt daher erst im Zuge der Vorstellung weiterführender Konzepte in Abschnitt 12.3.2. Dort wird gezeigt, dass diese Komplotte nur eine geringe Zahl von Mitgliedern aufnehmen und sich langfristig nicht durchsetzen können. Im Folgenden gehen wir daher auf die Möglichkeit solcher Komplotte nicht weiter ein.

- *Einseitige Ausrichtung des Komplotts:* Das Problem gegenseitiger Komplotte tritt nicht auf, wenn im Komplott nur eine bestimmte Einheit unterstützt wird. In der manipulierten Version findet sich zu diesem Zweck unveränderlich durch Angabe der Identität vordefiniert, welche Einheit zu unterstützen ist. Dieser Ansatz ist für den Ersteller einer manipulierten Version besonders viel versprechend, da er seine eigene Einheit als Nutznießer des Komplotts verankern kann. Allerdings ist nicht zu erwarten, dass eine manipulierte Version, die individuell irrationales Verhalten zugunsten ihres Erstellers vorschreibt, von anderen Benutzern auch tatsächlich übernommen wird. Viel eher werden sich übernahmewillige Benutzer eine andere manipulierte Version beschaffen, die ihnen weniger oder keinen Nachteil bietet. Die Ersteller manipulierter Versionen stehen damit in Konkurrenz zueinander. Als Ergebnis dieser Konkurrenzsituation setzt sich diejenige manipulierte Version durch, die im Interesse der übernahmewilligen Benutzer handelt. Damit kann sich eine manipulierte Version, die ein einseitig ausgerichtetes Komplott bezweckt, nicht durchsetzen.

Es zeigt sich also, dass keine der beiden Möglichkeiten zur Bereitstellung individuell irrationaler Versionen gangbar ist. Als Folge davon erhalten wir, dass lediglich manipulierte Versionen, die individuell rational sind, verwendet werden.

Damit gestaltet sich das Modell der Einheiten wie folgt: Normative Einheiten verhalten sich entsprechend des Systementwurfs, während strategische Einheiten sich immer zu dem Verhalten entscheiden, das ihnen den größten Vorteil bietet.

5.3.4 Vergleich mit evolutionären Modellen von Einheiten

In den vorigen Abschnitten ist das Modell der Einheiten vorgestellt und untersucht worden, das dieser Arbeit zugrunde liegt. Es leitet sich aus den Eigenschaften des Campus-Szenarios, insbesondere der Autonomie der Teilnehmer, ab. In der Literatur wird im Gegensatz dazu des Öfteren eines der evolutionären Modelle von Einheiten angenommen [CCP98, BB02a].

Dieser Abschnitt beschäftigt sich damit, ob diese Annahme gerechtfertigt ist. Dazu wird das eigene, dem Campus-Szenario angemessene Modell der Einheiten mit dem evolutionären Modell verglichen, das gemäß Abschnitt 4.1.3 den Eigenschaften der Kooperation im Campus-Szenario am nächsten kommt.

Gemeinsamkeiten. Beide Modelle haben einige Gemeinsamkeiten. Sie betreffen die folgenden Punkte:

- *Vor-Festlegung zu gewissem Verhalten in Strategien:* Im eigenen Modell ist das Verhalten einer Einheit durch die Version der auf ihrem Gerät verwendeten Systemsoftware bestimmt. Im evolutionären Modell kommt es auch zu einer solchen Vor-Festlegung (engl: pre-commitment), weil das Verhalten der Einheiten von ihrer jeweiligen Strategie definiert wird.

- *Entstehung neuer Strategien:* Neue Strategien entstehen im eigenen Modell aus der Erstellung einer neuen manipulierten Version der Systemsoftware. Eine solche Erstellung erfolgt typischerweise dadurch, dass bestimmte Teile der originalen Systemsoftware angepasst werden. Strategien entstehen in ähnlicher Weise im evolutionären Modell durch Mutation oder Rekombination.

- *Dynamik des Gesamtsystems durch eine Rückkopplung:* In beiden Modellen kommt es zu Entscheidungen auf individueller Ebene. Diese bestehen beim eigenen Modell in der Typ-

wahl der Benutzer, im evolutionären Modell in der Übernahme erfolgreicher Strategien. Die individuellen Entscheidungen stehen in einem beidseitigen Zusammenhang mit den Systemeigenschaften. Durch ihre Rückkopplung folgt in beiden Modellen eine Dynamik der Populationsstruktur, die zur Evolution des Gesamtsystems führt.

Die Existenz dieser Gemeinsamkeiten gibt eine Erklärung dafür, dass in der Literatur des Öfteren das evolutionäre Modell von Einheiten als angemessen für Informationssysteme wie im Campus-Szenario betrachtet wird. Allerdings gibt es trotz dieser Gemeinsamkeiten mehrere Punkte, in denen die beiden Modelle grundsätzlich verschieden veranlagt sind.

Unterschiede. Die beiden Modelle unterscheiden sich unter anderem darin, welche Populationsstruktur sie initial annehmen. Im eigenen Modell ist festgehalten, dass initial nur normative Einheiten zu erwarten sind. Dies gilt bis zu dem Zeitpunkt, an dem eine manipulierte Version der Systemsoftware erstmalig erstellt wird. Im Gegensatz dazu macht das evolutionäre Modell keine Aussagen über die initiale Populationsstruktur.

Die restlichen Unterschiede zwischen den beiden Modellen betreffen die Eigenheiten der individuellen Entscheidungen. Hierbei sind folgende Punkte verschieden in den Modellen festgelegt:

- *Urheber der individuellen Entscheidung:* Im eigenen Modell entscheidet der Benutzer durch seine Typwahl. Es besteht daher eine feste Bindung zwischen dem Benutzer und seiner Einheit. Insbesondere ist der Benutzer frei zu entscheiden, aus dem System auszutreten. Ein gegensätzliches Bild bietet sich beim evolutionären Modell. Dort bestimmt die Replikationsdynamik, welche Einheiten welche Strategien übernehmen. Insbesondere kann sich die Gesamtzahl der Einheiten erhöhen oder vermindern. Es fehlt hier also eine feste Bindung zum menschlichen Benutzer als den Urheber der individuellen Entscheidung.

- *Hindernisse bei der Revision der individuellen Entscheidung:* Im evolutionären Modell wird eine Strategie, die von anderen Einheiten erfolgreich verfolgt wurde, mit einer gewissen Wahrscheinlichkeit übernommen. Dies unterscheidet sich von der Revision der Typwahl im eigenen Modell. Dort bestimmt der Benutzer durch utilitaristische Überlegungen, welche Version der Systemsoftware er verwendet. Durch den Wechsel zu einer manipulierten Version entstehen Manipulationskosten. Es besteht daher im Gegensatz zum evolutionären Modell ein Hindernis dafür, erfolgreiche Strategien anderer zu übernehmen.

Es zeigt sich also, dass die Typwahl eines menschlichen Benutzers sich grundsätzlich von der Replikationsdynamik unterscheidet.

Fazit. Trotz ihrer Gemeinsamkeiten entsprechen sich das eigene und das evolutionäre Modell nicht gegenseitig. Daher ist keine direkte Übertragung der Ergebnisse der evolutionären Spieltheorie möglich. Diese Feststellung trifft umso mehr zu, als das evolutionäre Modell, das gemäß Abschnitt 4.1.3 dem Campus-Szenario am nächsten kommt und daher der Untersuchung dieses Abschnitts zugrunde liegt, in der Literatur der evolutionären Spieltheorie als Spezialfall kaum untersucht wurde.

Eine weitere Schlussfolgerung bietet sich für den Systementwurf an. Das eigene Modell weist dem Entwurf der originalen Systemsoftware eine hohe Wichtigkeit zu. Dies liegt an zwei Besonderheiten des eigenen Modells im Vergleich zum evolutionären Modell: **(1)** Der Systementwurf bestimmt die initiale Populationsstruktur und damit die Eigenschaften, die das Informationssystems anfangs besitzt. **(2)** Aufgrund der Hindernisse zur Manipulation tendiert der Benutzer zur

Verwendung der originalen Systemsoftware, solange der Systementwurf für hinreichend geringe Normativitätskosten sorgt.

5.4 Methodik des Systementwurfs

Das Modell der Einheiten gibt Aufschluss darüber, wie sich Einheiten verhalten und sich das Informationssystem entwickelt. Die Aussagen des Modells stehen mit den Eigenschaften der originalen Systemsoftware in einem engen Zusammenhang. Diese beeinflusst nachhaltig die Typwahl der Einheiten, weil die Höhe der Normativitätskosten und zum Teil auch der Manipulationskosten von ihr abhängen. Das Modell macht aber seiner Natur nach keine Aussagen darüber, wie die originale Systemsoftware zu entwerfen ist.

Dieser Abschnitt befasst sich daher mit der Methodik des Systementwurfs. Die Ausrichtung des Entwurfs wird in Abschnitt 5.4.1 diskutiert. Abschnitt 5.4.2 untersucht, inwiefern die Verhaltensvorschriften, die im Entwurf festgelegt werden, Normen für die Einheiten darstellen. Der Definition der Normen fällt eine entscheidende Rolle beim Systementwurf zu. Abschnitt 5.4.3 stellt die Prinzipien vor, die dabei zu berücksichtigen sind, um die Normen selbstdurchsetzend zu machen.

5.4.1 Ausrichtung

Der Systementwurf zielt letztlich auf die Bestimmung der Systemeigenschaften. Aus Abschnitt 5.1 wurde aber ersichtlich, dass diese aufgrund der Autonomie der Teilnehmer nicht direkt bestimmt werden können. Die Rolle des Systementwerfers beschränkt sich darauf, den Benutzern eine Systemsoftware vorzuschlagen und zur Verfügung zu stellen. Die Benutzer nehmen diesen Vorschlag in Abhängigkeit von den Eigenheiten der originalen Systemsoftware an. Beim Systementwurf sind daher nicht nur wünschenswerte Eigenschaften des Gesamtsystems einzubeziehen. Darüber hinaus müssen die Eigenheiten der Systemsoftware im Hinblick auf die Typwahl der Benutzer bestimmt werden.

Dieser Abschnitt befasst sich mit der Frage, wie der Systementwurf auszurichten ist, um diese beiden Gesichtspunkte zu berücksichtigen. Dazu wird zunächst das Ziel des Systementwurfs und die Kriterien zu dessen Erreichung konkret gefasst. Anschließend werden zwei Maximen für den Systementwurf abgeleitet und untersucht, inwiefern sie miteinander im Konflikt stehen.

Ziel und Kriterien für die Zielerreichung. Die Existenzfähigkeit des Informationssystems wird gemäß Abschnitt 5.3.2 von zwei Seiten bedroht. Einerseits können sich Benutzer zum Austritt aus dem System entscheiden. Andererseits sind die Benutzer in der Lage, manipulierte Versionen der Systemsoftware zu verwenden, was letztendlich zum Austritt anderer Benutzer aus dem System führt. Um diese beiden Bedrohungen zu begegnen, muss das Ziel des Systementwurfs sein, dass *so viel Einheiten wie möglich normativ* sind. Dies ist gleichbedeutend damit, dass sich so viel Benutzer wie möglich dazu entscheiden, mit der originalen Systemsoftware am System teilzunehmen.

Ob der Systementwerfer sein Ziel erreicht, ist anhand zweier Kriterien überprüfbar, die sich direkt aus den Faktoren der Typwahl ableiten. Der Grad der Zielerreichung lässt sich daran erkennen, wie hoch der Anteil der Benutzer ist, bei denen diese beiden Kriterien erfüllt sind:

1. *Teilnahme:* Die Benutzer nehmen am System teil, da ihr Individualnutzen, der aus der Teilnahme entsteht, positiv ist.

2. *Normativität:* Die Benutzer verwenden die originale Systemsoftware, da ihre Manipulationskosten die Normativitätskosten übersteigen.

Auf den ersten Blick erscheint, dass das Kriterium der Teilnahme alleine ausreicht, um die Existenzfähigkeit des Informationssystems und damit die Zielerreichung des Entwurfs zu sichern. Dass dem jedoch nicht so ist, zeigt die folgende Überlegung: Manipulierte Versionen der Systemsoftware unterscheiden sich in wesentlichen Punkten von der originalen Systemsoftware. Nur dann lohnt es sich für einen Benutzer, die Manipulationskosten auf sich zu nehmen und eine manipulierte Version zu verwenden. Wir müssen daher erwarten, dass strategische Einheiten vom Systementwurf abweichen. Ist eine Vielzahl der teilnehmenden Einheiten strategisch, so unterscheiden sich damit auch die Systemeigenschaften wesentlich von denen, die im Systementwurf vorgesehen sind. Dadurch ist die Teilnahme der Benutzer nicht mehr gewährleistet. Daraus wird ersichtlich, dass auch das zweite Kriterium, nämlich das der Normativität, benötigt wird. Es verhindert, dass das Informationssystem degeneriert und dadurch das Kriterium der Teilnahme angegriffen wird.

Ein weiterer Punkt der Zieldefinition, der einer genaueren Erklärung bedarf, ist, dass lediglich ein hoher Anteil normativer Einheiten gefordert wird. Warum wird nicht verlangt, dass alle Einheiten normativ sind? Die Antwort hierauf findet sich in der Heterogenität der Manipulationskosten. Wie in Abschnitt 5.3.1 gezeigt, ist davon auszugehen, dass einige wenige Benutzer sehr geringe oder keine Manipulationskosten besitzen. Diese inhärent manipulationsfreudigen Benutzer von der Verwendung der originalen Systemsoftware zu überzeugen, ist daher äußerst schwierig oder unmöglich. Des Weiteren ist es auch gar nicht notwendig, dass alle Einheiten normativ sind. Wenn bei einem hohen Anteil von Benutzern die beiden Kriterien erfüllt sind, so bedeutet dies unter anderem, dass die Normativitätskosten gering sind. Ansonsten wäre es zu einer weiteren Verbreitung manipulierter Versionen gekommen. Niedrige Normativitätskosten bedeuten, dass die inhärent manipulationsfreudigen Benutzer gegenüber diejenigen Benutzer, die die originale Systemsoftware benutzen, kaum einen Vorteil haben. Insofern ist es auch unter dem Gesichtspunkt der Fairness tragbar, dass nicht alle Einheiten normativ sind. Entscheidend für den Erfolg des Systementwurf ist einzig, dass die meisten Benutzer aus eigenem Interesse mit der originalen Systemsoftware am System teilnehmen. Eben dies wird in der Zieldefinition gefordert.

Maximen des Systementwurfs. Um die Kriterien der Zielerreichung zu erfüllen, müssen die drei Faktoren der Typwahl entsprechend beeinflusst werden. Der Individualnutzen entscheidet über das Kriterium der Teilnahme, während die Normativitätskosten und Manipulationskosten über das Kriterium der Normativität bestimmen. Der Systementwurf hat also dafür zu sorgen, dass der Individualnutzen und die Manipulationskosten maximiert und die Normativitätskosten minimiert werden.

Die Möglichkeiten der Einflussnahme des Systementwerfers auf die Manipulationskosten der menschlichen Benutzer ist eher beschränkt. Dies liegt daran, dass die Hindernisse zur Manipulation nur zum Teil vom Systementwurf abhängen. Im Wesentlichen beschränkt sich der Aufgabenbereich des Systementwerfers im Hinblick auf die Manipulationskosten daher auf zwei Punkte:

- Vor Ausgabe der Systemsoftware sind auf ihr die Techniken der Code-Verschleierung anzuwenden. Damit wird sichergestellt, dass es ein technisches Hindernis zur Erstellung manipulierter Versionen gibt.
- Die Ausgabe der Systemsoftware und die Vergabe der zertifizierten Identitäten erfolgt nur unter einer Lizenz, die entsprechend Abschnitt 5.2.2 die Verwendung manipulierter Versionen vertragsrechtlich angreifbar macht.

Beide Punkte lassen sich am Ende des Systementwurfs unabhängig von dessen Eigenheiten berücksichtigen. Daher widmen wir uns im Systementwurf den verbleibenden Aufgaben, nämlich der Maximierung des Individualnutzens und der Minimierung der Normativitätskosten. Sie lassen sich in Form von zwei Maximen formulieren:

- *1. Maxime:* Normative Einheiten verhalten sich *hinreichend kooperativ*. Die Gutartigkeit der normativen Einheiten wird benötigt, damit im Informationssystem kooperatives Verhalten vorherrscht. Diese wünschenswerte Systemeigenschaft ermöglicht, dass für die Benutzer ein Mehrwert aus der Teilnahme am System entsteht. Diese Maxime sorgt also für die Maximierung des Individualnutzens.

- *2. Maxime:* Normative Einheiten verhalten sich *hinreichend vorteilhaft*. Dies ist erforderlich, um zu verhindern, dass sich die Prinzipale normativer Einheiten zur Manipulation entscheiden. Die Minimierung der Normativitätskosten ist also das Ziel dieser Maxime.

Konflikt zwischen den Maximen. Die Maximen des Systementwurfs fordern, dass hinreichend kooperatives und vorteilhaftes Verhalten für die normativen Einheiten vorgesehen wird. Die Befolgung jeweils einer Maxime ist auf triviale Weise möglich:

- *Altruistischer Systementwurf:* Eine Einheit verhält sich altruistisch, wenn sie sich unabhängig von etwaigen zukünftigen Gegenleistungen oder Erwiderungen jederzeit zur Kooperation bereit zeigt. Wenn der Systementwurf für die normativen Einheiten altruistisches Verhalten vorsieht, so ist daher das Verhalten normativer Einheiten so kooperativ wie irgend möglich. Allerdings verletzt ein solcher altruistischer Systementwurf die zweite Maxime, da sich in der Anwesenheit von Altruisten das Betrugsverhalten strategischer Einheiten besonders lohnt und damit das Verhalten normativer Einheiten äußerst nachteilig ist.

- *Opportunistischer Systementwurf:* Die zweite Maxime fordert, dass der menschliche Prinzipal einer normativen Einheit keinen Vorteil darin sieht, eine manipulierte Version der Systemsoftware zu verwenden. Dies ist dann der Fall, wenn im Systementwurf vorgesehen wird, dass auch normative Einheiten jede Gelegenheit zu vorteilhaftem Betrugsverhalten ausnutzen. Ein solcher opportunistischer Systementwurf verletzt jedoch die Forderung der ersten Maxime, da sich dann normative Einheiten nur sehr eingeschränkt oder gar nicht kooperativ verhalten würden.

Die Betrachtung zeigt, dass jeweils eine der beiden Maximen durch einen trivialen Systementwurf verfolgt werden kann. Allerdings erscheint es so, dass die Erfüllung einer Maxime die Verletzung der anderen verursacht. Die Schwierigkeit des Systementwurfs liegt also darin, einen Ausgleich zwischen den beiden Maximen zu finden.

Die Abbildungen 5.6 und 5.7 stellen den Zielkonflikt zwischen den beiden Maximen auf schematische Weise dar. In der Vertikalen ist der Grad der Einhaltung der ersten Maxime aufgetragen. Das im Systementwurf vorgesehene Verhalten kann dabei zwischen den Extremen von altruistischem Verhalten und Betrugsverhalten in jeder Transaktion liegen. Die horizontale Achse zeigt, wie sehr die zweite Maxime berücksichtigt wird. Je nach Systementwurf sind die normativen Einheiten zu einem unterschiedlichen Grad den strategischen Einheiten gegenüber benachteiligt. Die zwei markierten Punkte entsprechen den beiden oben genannten Extremen des Systementwurfs, nämlich dem altruistischen und opportunistischen Systementwurf. Die Möglichkeiten des Systementwurfs liegen zwischen diesen beiden Punkten. Dies wird durch die gepunktete Linie angezeigt, deren Verlauf beispielhaft ist.

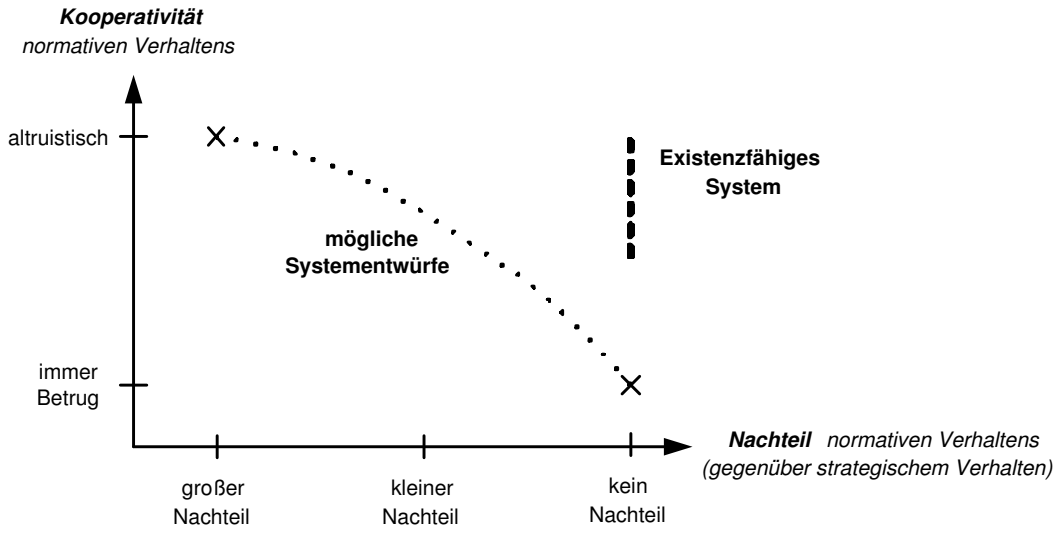

Abbildung 5.6: Zielkonflikt zwischen den Maximen des Systementwurfs ohne Berücksichtigung der Manipulationskosten

Nur wenn das Verhalten normativer Einheiten hinreichend kooperativ und vorteilhaft ist, ist das Informationssystem existenzfähig und der Systementwurf erfolgreich. Auf die Frage, was jeweils als hinreichend gilt, gehen die beiden Abbildungen auf unterschiedliche Weise ein. Da sie den Sachverhalt schematisch darstellen, machen sie keine quantitativen Aussagen darüber, was als hinreichend gilt. Dies wird in der Evaluation dieser Arbeit in Teil III untersucht. Der Zweck dieser beiden Abbildungen liegt vielmehr darin, die Intuition, die hinter den weiteren Schritten des Systementwurfs liegt, deutlich zu machen:

- *Ohne Manipulationskosten:* Wenn es keine Hindernisse zur Manipulation gibt, darf das Verhalten normativer Einheiten keinen Nachteil gegenüber dem von strategischen Einheiten mit sich bringen. Dies zeigt in Abbildung 5.6 die gestrichelte Linie rechts oben an. Sie gibt an, wann das Informationssystem existenzfähig ist. Aus den vorigen Überlegungen zum opportunistischen Systementwurf erhalten wir, dass keiner der möglichen Systementwürfe zu einem existenzfähigen Informationssystem führt. In der Abbildung wird dies daraus ersichtlich, dass die gepunktete Linie nicht die gestrichelte Linie schneidet.

- *Mit Manipulationskosten:* Die Situation ändert sich, wenn es Hindernisse zur Manipulation gibt. In diesem Fall entscheidet sich ein Benutzer nur dann zur Verwendung manipulierter Versionen, wenn die Normativitätskosten seine Manipulationskosten übersteigen. Wie in Abschnitt 5.2 gezeigt, sind die Manipulationskosten nicht unerheblich. Daher ist wird ein Benutzer sich erst dann zur Manipulation entschließen, wenn das Verhalten normativer Einheiten mehr als nur einen kleinen Nachteil mit sich bringt. Als Folge davon ist die Forderung der zweiten Maxime nach hinreichender Vorteilhaftigkeit normativen Verhaltens nicht mehr so stark. Abbildung 5.7 zeigt den resultierenden Bereich, in dem beide Maximen erfüllt sind und das Informationssystem dadurch existenzfähig ist. Gemäß der Darstellung sind damit einige der möglichen Systementwürfe in der Lage, beide Maximen zu erfüllen. Die Frage, ob dies für einen Systementwurf tatsächlich zutrifft, kann jedoch nur in einer quantitativen Evaluation des Systementwurfs erfolgen, wie sie in Teil III dieser Arbeit durchgeführt wird.

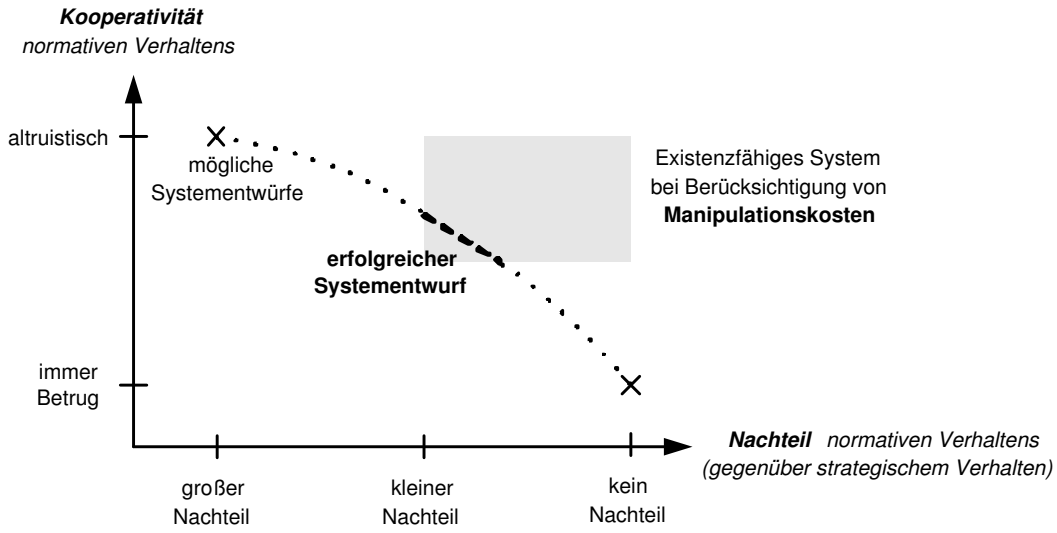

Abbildung 5.7: Zielkonflikt zwischen den Maximen des Systementwurfs unter Berücksichtigung der Manipulationskosten

Wir fassen also zusammen, dass ein Kompromiss zwischen den beiden Maximen des Systementwurfs nur dadurch ermöglicht wird, dass die Manipulationskosten nicht unerheblich sind. Ob jedoch tatsächlich ein solcher Kompromiss gefunden wird, hängt von der Güte des Systementwurfs ab. Daher muss die Methodik des Systementwurfs darauf eingehen, wie die Maximen umgesetzt werden.

5.4.2 Vorschriften und Normen

Die beiden Maximen des vorigen Abschnitts geben dem Systementwurf seine Ausrichtung. Es bleibt aber noch zu klären, auf welche Weise der Systementwurf diese Maximen verfolgen kann.

In diesem Abschnitt zeichnen wir ein differenziertes Bild vom Systementwurf. Er bedient sich zweier verschiedener Elemente, nämlich Verhaltensvorschriften und Verhaltensregeln. Letztere stellen Normen dar, weil sich die Einheiten anhand dieser Regeln gegenseitig kontrollieren. Im Folgenden werden diese Elemente des Systementwurfs vorgestellt.

Verhaltensvorschriften. Im Systementwurf wird die Funktionsweise der originalen Systemsoftware festgelegt. Diese bestimmt wiederum, wie sich normative Einheiten im Informationssystem verhalten. Der Systementwerfer ist daher in der Lage vorzuschreiben, wie sich normative Einheiten verhalten. Der Systementwurf besteht demnach aus einer Reihe von *Verhaltensvorschriften*.

Diese Verhaltensvorschriften umfassen alle Aspekte der Kommunikation und Kooperation im Informationssystem. Darunter fallen nicht nur die Protokolle, die bei der Abwicklung einer Transaktion oder bei der verteilten Vertrauensbildung anzuwenden sind. Darüber hinaus legen die Verhaltensvorschriften auch fest, wie normative Einheiten Entscheidungen, zum Beispiel bezüglich der Teilnahme an einer Transaktion und des Verhaltens darin, treffen. Der Systementwurf lässt sich also dadurch beschrieben, dass die jeweiligen Verhaltensvorschriften (oder kurz Vorschriften) für die normative Einheiten genannt werden.

Normen als Verhaltensregeln. Die verteilte Vertrauensbildung ist darauf aus, eine selbstorganisierende soziale Kontrolle zwischen den Einheiten einzurichten. Die Kontrolle bezieht sich nur auf einige wenige Verhaltensvorschriften, die im Folgenden *Verhaltensregeln* genannt werden. Die Beschränkung der Zahl der Verhaltensregeln ergibt sich daraus, dass die Kontrolle der meisten Vorschriften mit einem sich verbietenden Aufwand verbunden ist oder erst gar nicht möglich ist. Ein Beispiel hierfür sind die Vorschriften darüber, wie eine normative Einheit ihren Glauben über andere Einheiten bildet. Die Befolgung dieser Vorschrift kann nur indirekt über die Vertrauensentscheidungen der Einheit beobachtet werden und erfordert zudem, dass bekannt ist, über welches Wissen die Einheit verfügt. Es gibt jedoch auch Vorschriften, die sich als Verhaltensregeln anbieten. Beispielsweise ist eine Vorschrift über das Transaktionsverhalten einer Einheit ein viel versprechender Kandidat hierfür. Dies liegt daran, dass das Transaktionsverhalten einer Einheit zum Teil wahrnehmbar ist und für nachfolgende Transaktionen von hoher Wichtigkeit ist.

Das Herausstellen einzelner Verhaltensregeln erweitert den Rahmen des Systementwurfs. Er besteht nicht nur in der Angabe einer Reihe von Verhaltensvorschriften. Darüber hinaus legt er auch fest, welche dieser Vorschriften Verhaltensregeln sind und durch welche weiteren Vorschriften diese Regeln kontrolliert werden.

Im Gegensatz zu einfachen Vorschriften geben Verhaltensregeln an, welches Verhalten der Systementwerfer bei jeder Einheit durchsetzen will. In dieser Hinsicht handelt es sich bei einer Verhaltensregel um eine *Norm*, deren Befolgung auch von strategischen Einheiten erwartet wird. Dieser Begriff der Norm entspricht dem Konzept, das in der Soziologie mit *r-Norm* [Tuo95] bezeichnet wird: Eine Norm ist eine Beschränkung des Verhaltenrepertoires der Systemteilnehmer [CCP98]. Hingegen legen Vorschriften lediglich das Verhalten normativer Einheiten fest. Wie aus der Definition ersichtlich wird, sind Normen gleichzeitig auch Vorschriften, da sie als Verhaltensregeln von den normativen Einheiten eingehalten werden.

5.4.3 Selbstdurchsetzung des Entwurfs

Die beiden Maximen des Systementwurfs stehen miteinander in Konflikt, da kooperatives Verhalten für die jeweilige Einheit im Allgemeinen unvorteilhaft ist. Der Konflikt wird zwar zum Teil durch die Präsenz von Manipulationskosten gelöst, die einen gewissen Grad von Unvorteilhaftigkeit normativen Verhaltens erlauben. Jedoch ist auch im Entwurf selbst durch eine passende Definition der Vorschriften und Normen dafür zu sorgen, dass kooperatives Verhalten nur kaum oder gar nicht unvorteilhaft ist. In diesem Fall setzen sich die Vorschriften und Normen von selbst durch.

Dieser Abschnitt betrachtet die Aspekte einer solchen Selbstdurchsetzung (engl.: self-enforcement) des Entwurfs. Dazu wird sie zunächst mit der sozialen Kontrolle in Verbindung gebracht. Anschließend wird auf die Selbstdurchsetzung von Normen und Vorschriften jeweils eingegangen.

Selbstdurchsetzung und soziale Kontrolle. Um Betrugsverhalten effektiv einzudämmen, müssen sich die Einheiten untereinander kontrollieren. Diese soziale Kontrolle besteht darin, dass die Einheiten, deren Verhalten gegen die Normen verstößt, im Vergleich zu normeinhaltenden Einheiten bestraft werden. Dabei stellt sich die Frage, von wem diese Bestrafung durchführt wird. Aufgrund der Selbstorganisation der sozialen Kontrolle stehen hierfür lediglich die Einheiten selbst zur Verfügung. Soziale Kontrolle existiert daher nur dann, wenn sie von den Einheiten selbst ausgeübt wird.

Normative und strategische Einheiten unterscheiden sich grundsätzlich in ihrer Ausübung

der sozialen Kontrolle. Normative Einheiten folgen den Vorschriften und Normen des Systementwurfs und sind dadurch auf die Beteiligung an der sozialen Kontrolle vor-festgelegt (engl.: pre-committed). Anders sieht es bei den strategischen Einheiten aus. Diese üben nur dann soziale Kontrolle aus, wenn dies in ihrem Interesse liegt. Hierbei ergibt sich ein grundlegendes Problem des Systementwurfs: Die Bestrafung einer normverletzenden Einheit ist für die bestrafende Einheit nicht notwendigerweise von Vorteil. Um eine autonome Einheit effektiv zu bestrafen, müssen ihr Gelegenheiten zur Kooperation entzogen werden. Dies geschieht entweder direkt durch den Verzicht der bestrafenden Einheit auf weitere Kooperation mit der zu bestrafenden Einheit oder indirekt dadurch, dass die bestrafende Einheit eine entsprechende Empfehlung über die zu bestrafende Einheit ausstellt. In beiden Fällen entstehen der bestrafenden Einheit Kosten, so dass die Bestrafung einer normverletzenden Einheit zunächst keinen Vorteil für eine Einheit mit sich bringt. Ohne geeignete Maßnahmen im Systementwurf werden sich daher strategische Einheiten im Gegensatz zu den normativen Einheiten nicht an der sozialen Kontrolle beteiligen. Damit werden die Kosten für die soziale Kontrolle, die in [CCP98] *soziale Kosten* genannt werden, nicht gleich unter den Einheiten verteilt. Es folgt daraus eine weitere Erhöhung der Normativitätskosten.

Der Systementwurf darf sich daher nicht nur darauf beschränken, wünschenswertes Verhalten in der Definition der Normen festzuhalten. Darüber hinaus muss den strategischen Einheiten ein Anreiz gegeben werden, sich ebenfalls an der sozialen Kontrolle zu beteiligen. Ein trivialer Ansatz wäre hierbei, die Ausübung der sozialen Kontrolle selbst als Norm im Systementwurf vorzusehen. Das würde bedeuten, dass die Kontrolle der Befolgung der Normen in weiteren Normen vorgeschrieben wird. Offensichtlich löst dieser Ansatz das Problem nicht, er verschiebt es nur. Der Systementwurf muss also in seiner Definition der Normen und Vorschriften in der Hinsicht *abgeschlossen* sein, dass die Ausübung sozialer Kontrolle keine zusätzlichen Normen und Vorschriften benötigt.

Wie kann der Systementwurf eine solche Abgeschlossenheit erreichen? Für die strategischen Einheiten muss die Ausübung der sozialen Kontrolle einen direkten Vorteil erbringen, ohne dass diese Ausübung in einer Norm festgeschrieben und von den normativen Einheiten kontrolliert wird. Konkret bedeutet diese Forderung, dass eine strategische Einheit bei Beobachtung von normverletzendem Verhalten durch eine Einheit zur Bestrafung derselben bereit ist. Wenn wir zudem berücksichtigen, dass Normen nur durch strategische Einheiten verletzt werden können, erhalten wir folgendes Entwurfsprinzip:

Entwurfsprinzip 1: Paradoxie strategischen Verhaltens
Jede strategische Einheit verhält sich normativen Einheiten gegenüber *kooperativer* als strategischen Einheiten gegenüber.

Das im Entwurfsprinzip geforderte Verhalten ist insofern paradox, als gefordert wird, dass sich strategische Einheiten bevorzugt gegenseitig betrügen. Die Grundlage für die Umsetzung dieses Entwurfsprinzips ist die Beobachtung aus Abschnitt 5.3.3, dass strategische Einheiten sich individuell rational verhalten.

Wie ist dieses Entwurfsprinzip im Systementwurf zu berücksichtigen? Die Ausübung der sozialen Kontrolle durch die normativen Einheiten ist in der Definition der Vorschriften und Normen festgelegt. Wenn die Einhaltung dieser Vorschriften und Normen auch für die strategischen Einheiten von Vorteil ist, so ergibt sich dadurch automatisch, dass sich die strategischen Einheiten an der sozialen Kontrolle beteiligen. Der Systementwurf muss daher darauf achten, dass die Vorschriften und Normen so selbstdurchsetzend wie möglich sind. In den nachfolgenden Paragraphen besprechen wir diesen Gesichtspunkt separat für Normen und Vorschriften.

Selbstdurchsetzung von Normen. Damit sich Normen im Informationssystem durchsetzen, müssen sie nicht nur von den normativen Einheiten sondern auch von den strategischen Einheiten befolgt werden. Es stellt sich daher die Frage, wie die an sich nachteilige Einhaltung von Normen im Interesse einer strategischen Einheiten sein kann. Als Hilfsmittel steht hierfür die soziale Kontrolle zur Verfügung, die für die Überwachung Norm-bezogenen Verhaltens sorgt. Wie bereits erwähnt, ist der Entzug von Kooperationsmöglichkeiten das einzige Mittel, in einem selbstorganisierenden Informationssystem Einheiten zu bestrafen. Diese Bestrafung widerfährt eine Einheit, wenn sie durch wiederholtes normverletzendes Verhalten als nicht normativ erscheint. Daher erhalten wir folgendes Entwurfsprinzip:

Entwurfsprinzip 2:
Strategische Einheiten wollen als normative Einheiten wahrgenommen werden, um *nicht* durch den Entzug von Kooperationsmöglichkeiten *bestraft* zu werden.

Strategische Einheiten richten ihr Verhalten also danach aus, dass sie als normativ wahrgenommen werden. Aus der Sicht des Systementwerfers reicht aber dieses Entwurfsprinzip alleine nicht aus. Zusätzlich muss er dafür sorgen, dass Norm-bezogenes Verhalten auch tatsächlich von anderen Einheiten wahrgenommen wird. Nur dann entsteht für die strategischen Einheiten ein Anreiz, die Normen einzuhalten. Diese Überlegung wird in einem weiteren Entwurfsprinzip festgehalten:

Entwurfsprinzip 3:
Die einzige Möglichkeit für eine Einheit, als normativ wahrgenommen zu werden, besteht darin, die Normen einzuhalten.

Aus der Sicht der strategischen Einheiten bietet also die Einhaltung von Normen eine Gelegenheit dazu, die eigene Normativität den anderen Einheiten gegenüber zu *signalisieren*. Diese Situation entspricht der des Ladenketten-Paradoxons aus Abschnitt 4.1.2. Dieser Zusammenhang stellt sich wie folgt dar: Im Ladenketten-Paradox entscheidet sich der Monopolist zu einem Verhalten, das kurzfristig gesehen irrational ist. Dadurch täuscht er vor, von dem Typ von Monopolist zu sein, der auf die Bekämpfung seiner Konkurrenten vor-festgelegt ist. Der Monopolist verspricht sich von dieser Täuschung längerfristig einen höheren Nutzen, da seine Position durch seine vermeintliche Vorfestlegung nicht mehr angegriffen wird. Ebenso verhält es sich bei strategischen Einheiten. Sie entscheiden sich zum Einhalten der Normen, um sich als Einheiten normativen Typs auszugeben, die zu normativen Verhalten vor-festgelegt sind. Dafür nehmen sie in Kauf, dass die Einhaltung der Normen vorteilhaftes Betrugsverhalten ausschließt und damit kurzfristig gesehen von Nachteil ist. Das Ziel der Täuschung ist auch hier ein höherer Nutzen zu einem späteren Zeitpunkt, der aus der Verfügbarkeit späterer Kooperationsmöglichkeiten entsteht.

Wir halten zusammenfassend fest, dass strategische Einheiten sich an die Normen halten, um als normativ wahrgenommen zu werden. Als Voraussetzung muss der Systementwurf dafür sorgen, dass die beiden Entwurfsprinzipien umgesetzt werden. Konkret heißt dies, dass in den Vorschriften folgendes vorgesehen ist:

- Normativ erscheinende Einheiten werden für die Kooperation bevorzugt.
- Norm-bezogenes Verhalten wird durch entsprechende Protokolle und Verfahren so wahrnehmbar wie möglich gemacht.

Wenn beide Punkte in den Vorschriften berücksichtigt werden, üben die normativen Einheiten die notwendige soziale Kontrolle aus. Um die sozialen Kosten gerecht zwischen allen Einheiten aufzuteilen, müssen sich jedoch auch die strategischen Einheiten an der sozialen Kontrolle beteiligen. Es bleibt daher noch zu untersuchen, wie der Systementwurf erreichen kann, dass die entsprechenden Vorschriften der sozialen Kontrolle auch von den strategischen Einheiten befolgt werden.

Selbstdurchsetzung von Vorschriften. Um die Selbstdurchsetzung von Normen zu erreichen, muss der Systementwurf Vorschriften für die Bevorzugung normativ erscheinender Einheiten und die Wahrnehmbarkeit Norm-bezogenen Verhaltens enthalten. Diese Vorschriften sind dann selbstdurchsetzend, wenn auch strategische Einheiten sich an sie halten. Im Folgenden untersuchen wir also für diese beiden Punkte, warum die Befolgung der Vorschriften auch im Sinne der strategischen Einheiten ist.

Der erste Punkt befasst sich damit, wie strategische Einheiten ihre Transaktionspartner wählen. Erscheint ein potentieller Transaktionspartner als normativ, so ist eine Transaktion mit ihm vorteilhaft, da Betrugsverhalten von ihm nicht erwartet wird. Diese Überlegung gilt unabhängig davon, ob die strategische Einheit plant, in der Transaktion zu kooperieren oder den normativ erscheinenden Transaktionspartner zu betrügen. Anders sieht die Situation aus, wenn der potentielle Transaktionspartner strategisch erscheint. Da er sich zu Betrugsverhalten entschließen kann, verspricht sich eine Einheit von einer Transaktion mit ihm weniger oder gar keinen Nutzen. Als Folge davon wählt eine rational handelnde Einheit immer diejenige Einheit als Transaktionspartner, die am normativsten erscheint. Dies gilt also insbesondere auch für die Wahl von Transaktionspartnern durch strategische Einheiten. Damit ist gewährleistet, dass wie gefordert normativ erscheinende Einheiten für die Kooperation bevorzugt werden.

Der zweite Punkt fordert, dass auch strategische Einheiten die Protokolle und Verfahren des Systementwurfs einsetzen, die Norm-bezogenes Verhalten wahrnehmbar machen. Zu diesem Zweck muss für die Protokolle und Verfahren im Einzelfall nachgewiesen werden, dass diese Forderung von den Vorschriften des Systementwurfs erfüllt wird. Beispiele für einen solchen Nachweis finden sich bezüglich der Verwendung des Transaktionsprotokolls und dem Ausstellen negativer Empfehlungen in Abschnitt 7.6.2. Ein abschließender Nachweis über die Menge aller Vorschriften hinweg ist jedoch nur in der Evaluation des Systementwurfs möglich, der in Teil III dieser Arbeit simulativ durchgeführt wird.

5.5 Zusammenfassung

Dieses Kapitel hat sich mit dem Gegensatz zwischen Systementwurf und Autonomie befasst und mit den Folgen, die aus ihm entstehen.

Jeder menschliche Benutzer entscheidet *autonom* darüber, ob und mit welcher Software sein Gerät am Informationssystem teilnimmt. Bei dieser Entscheidung kann er zwischen drei Alternativen wählen, die für jeweils einen Typ von Einheit stehen, nämlich normativ, strategisch und nicht-teilnehmend. Der Vorteil normativer Einheiten gegenüber strategischer Einheiten ergibt sich daraus, dass der menschliche Benutzer nicht die Hindernisse für die Verwendung manipulierter Versionen der Systemsoftware bewältigen muss. Eine Analyse hat gezeigt, dass es sowohl für die Erstellung als auch für das Übernehmen einer manipulierten Version Hindernisse technischer und rechtlicher Art gibt. Daraus entstehen nicht vernachlässigbare Manipulationskosten für die Benutzer. Diese sind mit den Normativitätskosten, die bei der Wahl zu einer normativen Einheit

auftreten, abzuwägen. Es besteht eine Rückkopplung zwischen der Typwahl der einzelnen Benutzer untereinander. In einer Analyse haben wir dargestellt, welche Dynamik sich daraus für die Populationsstruktur ergibt. Außerdem ist gezeigt worden, dass strategische Einheiten sich nicht gegenseitig unterstützen, auch wenn auf ihren jeweiligen Geräten dieselbe manipulierte Version verwendet wird. Das sich ergebende Modell der Einheiten wurde mit dem evolutionären Modell von Einheiten verglichen. Trotz einiger Gemeinsamkeiten bestehen einige substantielle Unterschiede.

Auf der anderen Seite steht der *Systementwurf*, der sich mit den Verfahren befasst, die bei der originalen Systemsoftware zum Einsatz kommen. Das eigentliche Ziel des Systementwurfs ist es, die Eigenschaften des Systems zu bestimmen. Jedoch kann der Systementwerfer nicht durchsetzen, dass seine Systemsoftware auch tatsächlich von den autonomen Benutzern verwendet wird. Die Autonomie steht also deswegen im Gegensatz zum Systementwurf, weil sie ihm die direkte Bestimmung der Systemeigenschaften verwehrt. Als Folge davon muss das Ziel des Systementwurfs sein, dass die Benutzer aus ihrem eigenen Interesse die originale Systemsoftware verwenden. Die Kriterien für die Zielerreichung leiten sich aus denen der Typwahl der Benutzer ab. Um diese Kriterien zu erfüllen, wurden die zwei Maximen des Systementwurfs vorgestellt, die hinreichend kooperatives und hinreichend vorteilhaftes Verhalten für normative Einheiten verlangen. Diese Maximen stehen zwar miteinander im Konflikt, können aber durch die Existenz von Manipulationskosten in Einklang gebracht werden. Eine Umsetzung finden die Maximen in den Verhaltensvorschriften und Normen des Systementwurfs. Die Definition der Vorschriften und Normen orientiert sich maßgeblich daran, dass sie sich im System durchsetzen, indem ihre Befolgung für die jeweilige Einheit von Vorteil ist.

Kapitel 6

不逆詐 不憶不信 抑亦先覺者 是賢乎

"Ist jemand, der weder mit der Falschheit der Anderen rechnet noch berücksichtigt, dass Versprechen nicht eingehalten werden können, oder jemand, der sich im Vorhinein über die Möglichkeiten des Betrugs bewusst ist, der wahre Weise?"
(Gespräche und Aussprüche des Konfuzius, 14.33)

Lokale Vertrauensbildung

Der eigene Ansatz zielt darauf, dass sich die autonomen Einheiten des Informationssystems auf selbstorganisierende Weise untereinander kontrollieren. Eine solche soziale Kontrolle wird durch den Einsatz einer verteilten Vertrauensbildung ermöglicht. Der Ausgangspunkt für die Vertrauensbildung besteht darin, dass jede Einheit durch Teilnahme an Transaktionen ihre eigenen Erfahrungen über das Verhalten Anderer macht. Basierend auf diesen Erfahrungen bildet sich jede Einheit ihren Glauben über den Typ anderer Einheiten und fällt ihre Vertrauensentscheidungen. Eine solche Vertrauensbildung ist insofern lokal, als in sie lediglich die eigenen Erfahrungen der jeweiligen Einheit einfließen.

Dieses Kapitel stellt den Entwurf der lokalen Vertrauensbildung vor. Abschnitt 6.1 führt in ihre Funktionsweise ein. Anschließend werden die Komponenten der lokalen Vertrauensbildung besprochen. Es handelt sich bei ihnen um die Transaktionen (Abschnitt 6.2), die Glaubensbildung (Abschnitt 6.3) und die Vertrauensentscheidungen (Abschnitt 6.4).

6.1 Einführung

Die Funktionsweise lokaler Vertrauensbildung ist bereits für die Besprechung der verwandten Arbeiten in Abschnitt 2.4.1 beschrieben worden. Der dort vorgestellte Kreislauf wird in diesem Abschnitt um einige Erweiterungen ergänzt. Darüber hinaus wird eine datenzentrische Sicht des Kreislaufs gegeben.

Erweiterter Kreislauf der lokalen Vertrauensbildung. Abbildung 6.1 zeigt den Kreislauf der lokalen Vertrauensbildung, der dem eigenen Ansatz zugrunde liegt. Er unterscheidet sich in mehreren Punkten von dem Kreislauf der verwandten Arbeiten, wie er in Abbildung 2.6 vorgestellt worden ist. Diese Punkte stellen Erweiterungen dar, deren Notwendigkeit sich aus den drei Anforderungen an die lokale Vertrauensbildung ergibt, die in Abschnitt 2.4.2 vorgestellt worden sind:

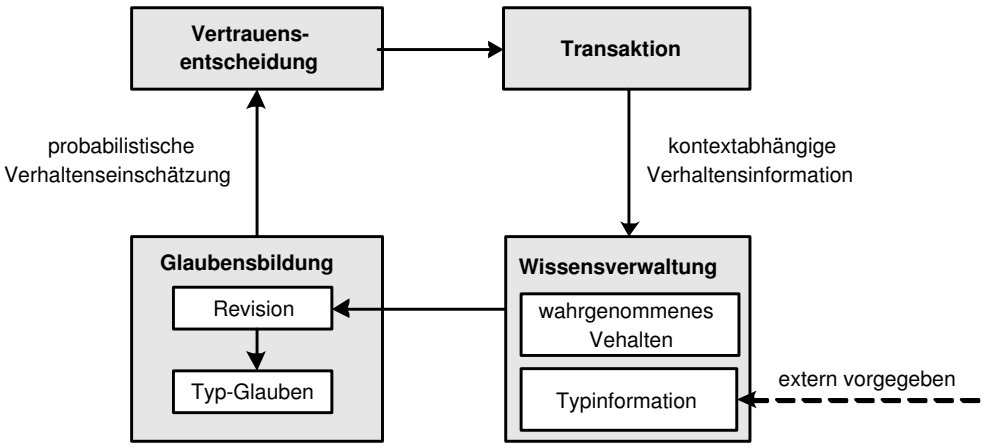

Abbildung 6.1: Der erweiterte Kreislauf der lokalen Vertrauensbildung

- *Berücksichtigung von Kontextinformation:* Das Verhalten, das ein Transaktionspartner zeigt, hängt vom Kontext der Transaktion, insbesondere ihrem Wert, ab. Die Wissensverwaltung legt daher nicht nur die Transaktionserfahrungen an sich ab, sondern auch in welchem Kontext sie sich ereignet haben. Damit wird ermöglicht, dass sich die Glaubensrevision auf Kontextinformation abstützt.

- *Berücksichtigung von Typinformation:* In einigen Fällen erhält eine Einheit Wissen über den Typ einer anderen Einheit. Bei einer rein lokalen Vertrauensbildung kann solches Wissen nur von außen, nämlich vom jeweiligen menschlichen Prinzipal, zugeführt werden. Dies liegt daran, dass Transaktionserfahrungen an sich kein abschließendes Urteil über den Typ des Transaktionspartners ermöglichen, da sich normative Einheiten unbeabsichtigterweise fehlverhalten und strategische Einheiten sich kooperativ verhalten können.

- *Quantitative Glaubensbildung:* Eine Einheit kann nur dann Vertrauensentscheidungen auf utilitaristische Weise treffen, wenn sie die Wahrscheinlichkeit einzuschätzen vermag, mit der ihr Transaktionspartner betrügen wird. Das Verfahren der Glaubensbildung und -revision muss probabilistisch fundiert sein, um eine solche Wahrscheinlichkeit von Betrugsverhalten quantifizieren zu können.

Ein weiterer Unterschied zu den verwandten Arbeiten besteht darin, dass die Glaubensbildung auf den Typ der Einheiten ausgerichtet ist. Dies wird notwendig, um diese drei Anforderungen gleichzeitig zu erfüllen.

Datenzentrische Sicht auf den Kreislauf. Eine alternative Sicht auf den Kreislauf der lokalen Vertrauensbildung zeigt Abbildung 6.2. Sie stellt die Daten und Informationen in den Vordergrund, die zwischen den Komponenten der lokalen Vertrauensbildung ausgetauscht werden. In dieser Hinsicht handelt es sich bei der Abbildung um eine datenzentrische Sicht auf den Kreislauf. Die Stationen des Kreislaufs sind wie folgt:

- *Glauben:* Jede Einheit besitzt ihren Glauben darüber, mit welcher Wahrscheinlichkeit eine andere Einheit in einer zukünftigen Transaktion betrügen wird. Basierend auf diesem Glauben werden Vertrauensentscheidungen getroffen.

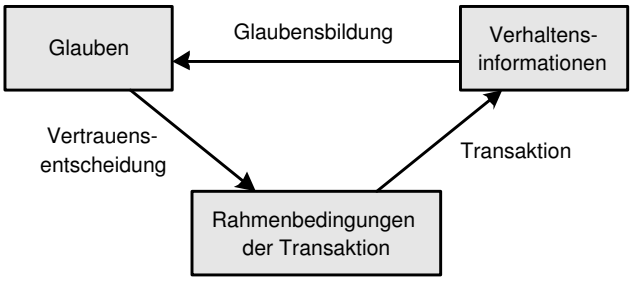

Abbildung 6.2: Datenzentrische Sicht auf den Kreislauf der lokalen Vertrauensbildung

- *Rahmenbedingungen der Transaktion:* Das Ergebnis einer Vertrauensentscheidung ist, mit welcher Einheit eine Transaktion eingegangen wird und welche Aktionen im Rahmen der Transaktion in welcher Reihenfolge ausgetauscht werden.
- *Verhaltensinformation:* Als Ergebnis einer Transaktion erhält eine Einheit Informationen darüber, wie sich der Transaktionspartner verhalten hat. Diese Verhaltensinformation wird der Glaubensbildung zugeführt.

Bei dieser Darstellung bleibt die Wissensverwaltung unberücksichtigt, da sie bei einer rein lokalen Vertrauensbildung keine andere Aufgaben hat, als die erhaltenen Verhaltensinformationen an die Glaubensbildung weiterzuleiten.

Das weitere Vorgehen dieses Kapitels ist entsprechend den Schritten dieses Kreislaufs gegliedert. Dabei gehen wir als erstes auf die Transaktionen ein. In den nachfolgenden Abschnitten werden die Glaubensbildung und die Vertrauensentscheidungen besprochen.

6.2 Transaktionen

Gemäß dem Systemmodell aus Abschnitt 1.2.2 besteht eine Transaktion darin, dass zwei Einheiten füreinander jeweils eine Aktion ausführen. Im Campus-Szenario besteht das Ausführen einer Aktion zum Beispiel aus dem Übermitteln einer Information, die vom Transaktionspartner benötigt wird.

In diesem Abschnitt beschäftigen wir uns mit den Vorschriften und Normen für die Teilnahme an Transaktionen. Der zeitliche Ablauf einer Transaktion wird im Zwei-Wege Transaktionsprotokoll festgelegt. Zudem wird die wichtigste Norm des Systementwurfs vorgestellt. Sie besagt, dass Einheiten ihre Transaktionspartner nicht betrügen sollen.

Zwei-Wege Transaktionsprotokoll. Bei einer Transaktion zwischen den beiden Einheiten A und B stellt sich die Frage, in welcher Reihenfolge die Einheiten ihre Aktionen ausführen. Wenn Einheit A ihre Aktion zuerst ausführt, so setzt sie sich der Gefahr aus, dass ihr Transaktionspartner B die Ausführung seiner Aktion unterlässt und damit Einheit A betrügt. Aus diesem Grund sprechen wir davon, dass die Einheit, die zuerst zieht, sich in der *Risiko-Position* befindet, während ihr Transaktionspartner eine *sichere Position* innehat.

Das sich daraus ergebende Transaktionsprotokoll ist in Abbildung 6.3 dargestellt. Im ersten Schritt führt Einheit A, die sich in der Risiko-Position befindet, ihre Aktion aus. Kommt Einheit B zum Schluss, dass diese Aktion korrekt ausgeführt wurde, so erwidert sie dies mit der Aktionsausführung ihrerseits in einem zweiten Schritt. Dieser Schritt wird wiederum von Einheit A

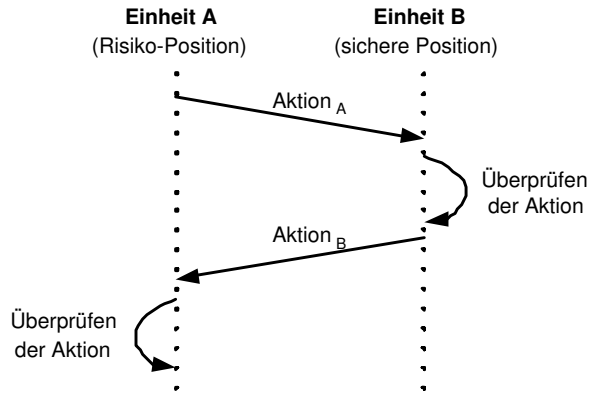

Abbildung 6.3: Das Zwei-Wege Transaktionsprotokoll

überprüft. Da das Protokoll aus zwei Schritten besteht, nennen wir es im Folgenden *Zwei-Wege Transaktionsprotokoll*. Es entspricht dem pessimistischen Austauschprotokoll aus Abschnitt 2.3.1 bis auf die Tatsache, dass das Ergebnis der Aktionsausführung dem jeweiligen Transaktionspartner direkt und damit ohne Vermittlung eines vertrauenswürdigen Dritten übermittelt wird. Auf einen solchen Dritten können wir aufgrund der Selbstorganisation nicht zurückgreifen.

Als Ergebnis der Transaktion stehen beiden Einheiten Beobachtungen darüber zur Verfügung, wie sich ihr jeweiliger Transaktionspartner verhalten hat. Eine solche Verhaltensinformation ist von einer der beiden folgenden Arten:

- *Kooperation:* Der eigene Transaktionspartner hat seine Aktion wie versprochen ausgeführt.
- *Betrug:* Die Aktion wurde vom Transaktionspartner nicht zur eigenen Zufriedenheit ausgeführt. Dies manifestiert sich in zwei Fällen die im Folgenden gleich behandelt werden: **(1)** Die Überprüfung ergibt, dass die Aktion des Transaktionspartners nicht korrekt ausgeführt worden ist. **(2)** Die Aktionsausführung des Transaktionspartners erfolgt nicht in einem vorgegebenen Zeitrahmen. In diesem Fall bricht die Einheit die Transaktion ab und bewertet dieses Ausbleiben der Aktionsausführung als Betrug.

Norm bezüglich des Transaktionsverhaltens. Das Transaktionsprotokoll macht deutlich, welche Aufgabe der Vertrauensbildung zufällt: Die Einheit in der sicheren Position ist davon abzubringen, durch Unterlassung ihrer Aktionsausführung zu betrügen. Um solches Betrugsverhalten einzudämmen, ist soziale Kontrolle zwischen den Einheiten unabdingbar. Beim Zwei-Wege Transaktionsprotokoll handelt es sich also nicht nur um eine Vorschrift, wie Transaktionen abzuwickeln sind. Darüber hinaus stellt diese Vorschrift auch eine Norm dar, deren Befolgung von Einheiten untereinander überwacht wird:

> **1. Norm:**
> Eine Einheit soll nicht beabsichtigen, im Laufe einer Transaktion zu betrügen.

Mit dieser Norm sind wir in der Lage, die erste Maxime des Systementwurfs einzuhalten: Das Verhalten der normative Einheiten ist in der Hinsicht hinreichend kooperativ, dass sie nie die Absicht haben zu betrügen.

Auf den ersten Blick erscheint es wenig intuitiv, dass die Definition der Norm nicht Betrugsverhalten an sich sondern nur die Absicht dazu verbietet. Unter Berücksichtigung unseres

Systemmodells löst sich aber dieser Widerspruch auf: Eine Einheit kann zwar beabsichtigen, ihre Aktion auszuführen. Ob sie jedoch dazu in der Lage ist, hängt von ihrer Umgebung ab. So führt vor allem ein Abbruch des Kommunikationskanals zwischen den beiden Transaktionspartnern dazu, dass die Aktionsausführung unbeabsichtigterweise fehlschlägt. Wenn also die Norm Betrugsverhalten an sich verbieten würde, käme es unter Umständen zur Verletzung der Norm selbst durch normative Einheiten. Daher orientiert sich die Definition der Norm an der Absicht der Transaktionspartner.

Eine weitere Erschwernis für die soziale Kontrolle ergibt sich daraus, dass das Transaktionsverhalten einer Einheit laut Systemmodell nur durch ihren jeweiligen Transaktionspartner beobachtet werden kann. Aus diesem Grunde nimmt die Mitteilung von Transaktionserfahrungen eine wichtige Rolle für die verteilte Vertrauensbildung ein. In diesem Kapitel steht allerdings zunächst im Vordergrund, wie eine Einheit ihre eigenen Transaktionserfahrungen bewertet.

6.3 Glaubensbildung

Jede Einheit wertet ihre Transaktionserfahrungen aus, um sich ihren Glauben über die anderen Einheiten zu bilden. In diesem Abschnitt entwerfen wir eine solche Glaubensbildung. Zu diesem Zweck führt Abschnitt 6.3.1 aus, dass die Glaubensbildung typorientiert sein muss, um die drei Anforderungen der lokalen Vertrauensbildung bewältigen zu können. Diese Einsicht wird in einem Glaubensmodell umgesetzt, das in Abschnitt 6.3.2 vorgestellt wird. Der Einsatz dieses Modells für die Glaubensrevision wird in Abschnitt 6.3.3 besprochen.

6.3.1 Typorientierung

Ziel der Glaubensbildung ist eine probabilistisch fundierte Einschätzung zukünftigen Verhaltens anderer Einheiten. Um dieses Ziel zu erreichen, stehen einer Einheit nicht nur ihre eigenen Transaktionserfahrungen zur Verfügung. Darüber hinaus sind gemäß der Anforderungen aus Abschnitt 2.4.2 der Kontext dieser Transaktionserfahrungen und etwaige Typinformationen zu berücksichtigen. Im Folgenden beschäftigen wir uns mit der Frage, welcher Ansatz für die Glaubensbildung gewählt werden muss, um diese Abbildung zwischen den zur Verfügung stehenden Informationen und der benötigten Einschätzung zukünftigen Verhaltens vornehmen zu können.

Die prinzipielle Vorgehensweise, die wir zu diesem Zweck verfolgen, ist in Abbildung 6.4 dargestellt. Der Glauben einer Einheit besteht demgemäß in der Einschätzung des Typs der anderen Einheiten. Dieser *Typglaube* wird auf die folgende Art und Weise gebildet und verwendet:

- *Revision:* Stehen neue Transaktionserfahrungen zur Verfügung, so führen sie zu einer Revision des Typglaubens. Diese Revision wird dabei abhängig vom Kontext der Transaktionserfahrungen durchgeführt.

- *Ableitung:* Für das Treffen von Vertrauensentscheidungen wird eine probabilistisch fundierte Einschätzung des zukünftigen Verhaltens Anderer benötigt. Diese wird nach Bedarf aus dem Typglauben abgeleitet.

Damit Revision und Ableitung auf diese Weise erfolgen können, ist ein Modell vonnöten, dass den Typ einer Einheit in Zusammenhang mit ihrem Verhalten bringt. Nur dann ist eine solche *typorientierte* Glaubensbildung durchführbar. Dieses Modell wird im nachfolgenden Abschnitt 6.3.2 entwickelt.

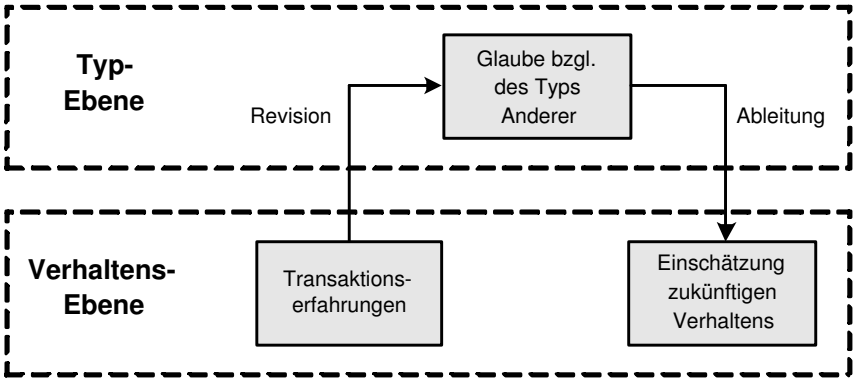

Abbildung 6.4: Funktionsweise der typorientierten Glaubensbildung

Auf den ersten Blick erscheint es wünschenswert, wie in den verwandten Arbeiten aus Abschnitt 2.4 die Glaubensbildung auf die Verhaltensebene zu verlegen. In einem solchen Fall entfällt die Notwendigkeit, einen Zusammenhang zwischen dem Typ und dem Verhalten einer Einheit herzustellen. Der Glaube einer Einheit besteht dann direkt in der Einschätzung zukünftigen Verhaltens Anderer. Es gibt allerdings mehrere Gründe dafür, die typorientierte Glaubensbildung einer solchen *verhaltensorientierten* vorzuziehen:

- *Kontext-Abhängigkeit von Verhalten:* Unabhängig davon, ob die Glaubensbildung sich am Typ oder am Verhalten einer Einheit orientiert, benötigt sie ein Modell davon, wie das Verhalten einer Einheit vom Kontext abhängt. Werden zum Beispiel Transaktionserfahrungen in einem Kontext gemacht, so stellt sich die Frage, wie der Glauben über das Verhalten in einem anderen Kontext revidiert werden muss. Die einzigen verwandten Arbeiten, die Kontextinformation berücksichtigen, nehmen dazu eine Korrelation zwischen verschiedenen Kontexten vor und führen einen separaten Verhaltensglauben für jeden Kontext. Dies führt zu einer äußert aufwändigen Glaubensverwaltung. In der typorientierten Glaubensbildung wird dieses Problem dadurch gelöst, dass kontextabhängiges Verhalten auf den Typ einer Einheit zurückgeführt wird. Die Korrelation zwischen verschiedenen Kontexten ist damit implizit in der Abbildung zwischen Typ und Verhalten einer Einheit integriert.

- *Einbeziehung von Typinformation:* Typinformation lässt sich in der typorientierten Glaubensbildung direkt berücksichtigen. Hingegen ist in der Literatur kein Verfahren der verhaltensorientierten Glaubensbildung bekannt, das mit Typinformation umgehen kann. Dieser Nachteil wiegt umso schwerer, als neben die externe Quelle von Typinformation in Kapitel 7 auch eine interne Quelle treten wird.

- *Typbasierte Verfahren im weiteren Entwurf:* Der Typglaube muss auch deswegen gebildet werden, weil er einigen Verfahren, die später in dieser Arbeit entwickelten werden, zugrunde liegt. Dies trifft sowohl auf das Treffen von Vertrauensentscheidungen in Abschnitt 6.4 als auch auf das Eingehen von sozialen Beziehungen in Kapitel 8 zu.

Wir halten also fest, dass die typorientierte Glaubensbildung der verhaltensorientierten Glaubensbildung vorzuziehen ist.

6.3.2 Glaubensmodell

Die typorientierte Glaubensbildung benötigt ein Modell, mit dem der Typ einer Einheit mit ihrem Transaktionsverhalten in Zusammenhang gebracht wird. Ein solches Glaubensmodell wird im Folgenden entwickelt.

Das TIB-Modell. Das Glaubensmodell muss zwei Sachverhalte berücksichtigen, um den Typ und das Verhalten einer Einheit in Beziehung zu setzen. Erstens können sowohl normative als auch strategische Einheiten die Absicht verfolgen, sich in der Transaktion kooperativ zu verhalten. Der Unterschied zwischen den beiden Typen von Einheiten liegt lediglich darin, ob eine gegenteilige Absicht, nämlich die zu Betrugsverhalten, möglich ist. Normative Einheiten beabsichtigen nie zu betrügen, da sie auf die Einhaltung der ersten Norm vor-festgelegt sind. Zweitens wird aus dem Systemmodell aus Abschnitt 1.2.2 ersichtlich, dass es trotz der Absicht zu kooperativem Verhalten unbeabsichtigterweise zu Betrugsverhalten kommen kann. Ist hingegen Betrug beabsichtigt, so kann sich daraus kein kooperatives Verhalten ergeben. Dies liegt daran, dass Betrug in der Unterlassung oder der Erbringung einer Fehlleistung besteht.

Diese Überlegungen werden in einem Glaubensmodell zusammengefasst, das wir im Folgenden Typ-Absicht-Verhalten (engl.: type-intention-behavior) Modell oder kurz *TIB-Modell* nennen. Es ist in Abbildung 6.5 dargestellt. Seine drei Ebenen entsprechen dem Typ, der Absicht und dem Verhalten einer Einheit. In der Besprechung dieser Ebenen gehen wir davon aus, dass sich Einheit X ihren Glauben über Einheit Y bildet:

- *Typglaube:* Einheit Y kann genau einem Typ angehören. Entweder ist sie normativ (Notation N_Y) oder strategisch (Notation S_Y oder $\overline{N_Y}$). Der Typglaube der Einheit X über Einheit Y besteht in ihrer Einschätzung der Wahrscheinlichkeit, dass es sich bei Einheit Y um eine normative Einheit handelt. Da es sich bei dieser Wahrscheinlichkeit um eine subjektive handelt, lässt sich dieser Typglaube also mit $p_X(N_Y)$ ausdrücken. Die Gleichsetzung des Glaubens mit einer subjektiven Wahrscheinlichkeit ergibt sich direkt aus der Besprechung des Abschnitts 4.2.

- *Absichtsglaube:* In einer Transaktion kann Einheit Y entweder die Absicht verfolgen zu kooperieren (engl.: cooperate) oder zu betrügen (engl.: defect). Dies notieren wir mit $C_Y^{(i)}$ beziehungsweise mit $D_Y^{(i)}$, wobei das hochgestellte i darauf hinweist, dass es sich um eine Absicht (engl.: intention) handelt. Der Glaube der Einheit X über die Absicht der Einheit Y wird Absichtsglaube genannt. Da er vom Kontext γ der Transaktion abhängt, wird er mit $p_X(C_Y^{(i)}|\gamma)$ angegeben.

- *Verhaltensglaube:* Entsprechend der Notation für die Absicht von Einheit Y kennzeichnen wir kooperatives Verhalten mit $C_Y^{(b)}$ und Betrugsverhalten mit $D_Y^{(b)}$, Das hochgestellte b bringt dabei zum Ausdruck, dass es sich um das Verhalten (engl.: behavior) der Einheit Y handelt. Analog zum Absichtsglauben wird der Verhaltensglaube mit $p_X(C_Y^{(b)}|\gamma)$ notiert.

Das TIB-Modell stellt eine Beziehung zwischen Typ und Verhalten einer Einheit über ihre Absicht her. Dies erfolgt in zwei Schritten:

- *Typ und Absicht:* Für eine normative Einheit gilt, dass sie immer die Absicht zu kooperativem Verhalten hat. Dies wird mit der bedingten Wahrscheinlichkeit $p(C_Y^{(i)}|N_Y) = 1$ zum Ausdruck gebracht. Auch strategische Einheiten können kooperatives Verhalten beabsichtigen. Die Wahrscheinlichkeit, mit der sie dies tun, hängt vom Transaktionskontext ab und

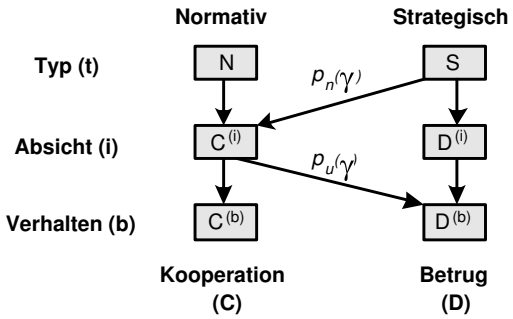

Abbildung 6.5: Zusammenhang zwischen Typ und Verhalten im TIB-Modell

wird im Folgenden mit $p_n(\gamma) = p(C_Y^{(i)}|S_Y, \gamma)$ bezeichnet. Dabei weist der Index n darauf hin, dass die Absicht zu kooperativem Verhalten von der Norm gefordert wird. Insgesamt erhalten wir also als Zusammenhang zwischen Typglauben und Absichtsglauben:

$$\begin{aligned} p_X(C_Y^{(i)}|\gamma) &= p_X(N_Y) + p_X(S_Y) \cdot p_n(\gamma) \\ p_X(D_Y^{(i)}|\gamma) &= p_X(S_Y) \cdot (1 - p_n(\gamma)) \end{aligned} \tag{6.1}$$

- *Absicht und Verhalten:* Die bedingte Wahrscheinlichkeit $p(C_Y^{(b)}|D_Y^{(i)}) = 0$ hält fest, dass es nur dann zu kooperativem Verhalten kommen kann, wenn eine entsprechende Absicht vorliegt. Kommt es zu Betrugsverhalten, so kann es allerdings auch unbeabsichtigt erfolgen. Die Wahrscheinlichkeit hierfür bezeichnen wir mit $p_u(\gamma) = p(D^{(b)}|C^{(i)}, \gamma)$, wobei der Index u darauf hinweist, dass Betrug unbeabsichtigt ist. Auch diese Wahrscheinlichkeit ist vom Transaktionskontext abhängig. Somit erhalten wir als Zusammenhang zwischen Absichtsglauben und Verhaltensglauben:

$$\begin{aligned} p_X(C_Y^{(b)}|\gamma) &= p_X(C_Y^{(i)}) \cdot (1 - p_u(\gamma)) \\ p_X(D_Y^{(b)}|\gamma) &= p_X(C_Y^{(i)}) \cdot p_u(\gamma) + p_X(D_Y^{(i)}) \end{aligned} \tag{6.2}$$

Aus der Darstellung wird ersichtlich, dass die drei Ebenen von Glauben im TIB-Modell sich auf verschiedene Arten von Unsicherheit beziehen. Der Typglaube steht lediglich unter dem Einfluss epistemischer Unsicherheit. Liegt einer Einheit zum Beispiel eine Typinformation bezüglich einer anderen Einheit vor, so ist sie sich über den Typ dieser Einheit nicht mehr unsicher. Anders verhält es sich mit dem Absichts- und Verhaltensglauben. Diese Ebenen des Glaubens leiten sich aus dem Typglauben unter Berücksichtigung der Wahrscheinlichkeit strategischer Einhaltung $p_n(\gamma)$ und unbeabsichtigten Betrugsverhaltens $p_u(\gamma)$ ab. Der Absichts- und Verhaltensglaube steht damit unter dem Einfluss sowohl der epistemischen als auch der stochastischen Unsicherheit.

In den zwei folgenden Paragraphen wenden wir uns der Parametrisierung des TIB-Modells zu. Hierfür ist zu untersuchen, wie die Wahrscheinlichkeit der strategischen Einhaltung und unbeabsichtigten Betrugsverhaltens einzuschätzen ist.

Einschätzung der strategischen Einhaltung. Unter welchen Umständen haben strategische Einheiten die Absicht, sich in einer Transaktion kooperativ zu verhalten? Zunächst einmal entgeht ihnen dadurch die Gelegenheit, Vorteile aus ihrem Betrugsverhalten zu ziehen. Andererseits beeinflusst eigenes kooperatives Verhalten den Transaktionspartner in seiner Glaubensbildung

derart, dass er eher zu zukünftigen Transaktionen bereit ist. Eine strategische Einheit wird sich also genau dann zur Einhaltung der Norm entscheiden, wenn die Möglichkeit längerfristiger Kooperation den kurzfristigen Betrugsvorteil in den Schatten stellt. Dabei schätzt sie den Nutzen aus längerfristiger Kooperation umso höher ein,

1. je länger ihr Prinzipal plant, am Informationssystem teilzunehmen,
2. je höher der erwartete Nutzen aus zukünftigen Transaktionen im Vergleich zu der aktuellen Transaktion ist[1],
3. je wahrscheinlicher es für sie ist, auf bisherige Transaktionspartner zu treffen,
4. *(nach Erweiterung zur verteilten Vertrauensbildung in den nachfolgenden Kapiteln)* je wahrscheinlicher eine Begegnung mit Einheiten ist, die aufgrund von Empfehlungen vom eigenen Transaktionsverhalten erfahren.

Das Problem dieser strategischen Faktoren liegt darin, dass außer der strategischen Einheit selbst keine andere Einheit sie einschätzen kann. Die einzige Ausnahme bildet hierbei der zweite Faktor: Der Transaktionspartner einer strategischen Einheit weiß, welche Aktion sie versprochen hat auszuführen. Bei den meisten Aktionen kann sich der Transaktionspartner daraus den relativen Wert der Transaktion für die strategische Einheit ableiten. Zum Beispiel ist im Campus-Szenario das Konvertieren von Annas Vertiefungslektüre mit Sicherheit aufwändiger als das Weiterleiten des Mensaplans.

Aus der Sicht einer Einheit, die die strategische Einhaltung einschätzen will, ergibt sich daher folgendes Bild: Sie kann ihre Einschätzung nur von einem der strategischen Faktoren, dem Transaktionswert, abhängig machen. Wenn dieser Wert im Vergleich zu durchschnittlichen Transaktionswerten sehr hoch ist, so ist von einer strategischen Einheit Betrugsverhalten zu erwarten. Dies liegt daran, dass für sie der Vorteil aus etwaigen zukünftigen Transaktionen geringer ist als der Betrugsvorteil in der aktuellen Transaktion. Umgekehrt ist bei einem sehr kleinen Transaktionswert kooperatives Verhalten zu erwarten, da der Betrugsvorteil im Vergleich zu späteren Transaktionen mit höherem Transaktionswert minimal ist. Für die Einschätzung von $p_n(\gamma)$ erhalten wir also zwei Bedingungen. Wenn die Funktion v für einen Kontext γ den Transaktionswert zurückgibt, lauten diese Bedingungen wie folgt:

$$\begin{aligned} \lim_{v(\gamma)\to 0} p_n(\gamma) &= 1 \\ \lim_{v(\gamma)\to \infty} p_n(\gamma) &= 0 \end{aligned} \tag{6.3}$$

Eine Einschätzung, die diese Bedingungen erfüllt, ist $p_n(\gamma) = \exp(-\kappa \cdot v(\gamma))$. Die Konstante κ gibt dabei ein, wie schnell mit steigendem Transaktionswert die Neigung zur Betrugsabsicht zunimmt. Die Quantifizierung dieser Konstante hängt maßgeblich von der Quantifizierung des Transaktionswerts und damit der Anwendungsdomäne ab. In Abschnitt A.2.1 des Anhangs findet sich ein Beispiel für eine solche Quantifizierung.

Wie ist mit einer Situation umzugehen, in der der Wert, den der Transaktionspartner der Transaktion beimisst, nicht genau abgeschätzt werden kann? Dieser Fall tritt dann ein, wenn die Kosten der Aktionsausführung nicht bekannt sind. Zum Beispiel verursacht das Konvertieren

[1] Unter dem Transaktionswert verstehen wir weiterhin, wie in Abschnitt 2.4.2 eingeführt, den Aufwand, den die eigene Aktionsausführung im Rahmen der Transaktion mit sich bringt.

von Annas Vertiefungslektüre für einen PDA einen höheren Aufwand als für einen ressourcenstarken Laptop. Die Funktion $v(\gamma)$ kann also nur eine mittlere Abschätzung des zu erwartenden Transaktionswerts sein. Im Extremfall lassen sich keine Informationen über den Transaktionswert ableiten. Dies würde bedeuten, dass die Einschätzung strategischer Einhaltung $p_n(\gamma)$ eine Konstante ist. Auch in diesem Fall ließe sich das TIB-Modell anwenden. Allerdings gäbe es damit keine Möglichkeit, kontextabhängiges Betrugsverhalten aufzuspüren und zu bestrafen. In Informationssystemen wie dem des Campus-Szenarios tritt dieses Problem jedoch nicht auf. Da wie gezeigt die Kosten der Aktionsausführung zumindest grob eingeschätzt werden können, ist eine Abschätzung des Transaktionswerts möglich.

Einschätzung unbeabsichtigten Betrugsverhaltens. Ursache für unbeabsichtigtes Betrugsverhalten sind Störungen in der Umgebung der Transaktion. In Ad-hoc-Netzen wie im Campus-Szenario hat hierbei der Abbruch des Kommunikationskanals eine vorherrschende Stellung. Die Einschätzung unbeabsichtigten Betrugsverhalten richtet sich daher nach der Wahrscheinlichkeit für solch einen Abbruch. Er ist umso wahrscheinlicher je länger eine Transaktion dauert. Dies liegt daran, dass die menschlichen Prinzipale bei einer länger andauernden Transaktion sich eher außer Reichweite bewegen. Die Dauer einer Transaktion hängt maßgeblich davon ab, welche Aktionen auszuführen sind, und ist damit eine Dimension des Transaktionskontextes. Im Campus-Szenario dauert die Übertragung eines Mensaplans zum Beispiel nur den Bruchteil einer Sekunde, während die Übertragung von Annas Vertiefungslektüre ungefähr eine Sekunde in Anspruch nimmt. Wir erhalten also zwei Bedingungen an die Einschätzung von $p_u(\gamma)$, wobei die Funktion d für einen Kontext γ die erwartete Dauer der Transaktion zurückgibt:

$$\begin{aligned} \lim_{d(\gamma)\to 0} p_u(\gamma) &= 0 \\ \lim_{d(\gamma)\to \infty} p_u(\gamma) &= 1 \end{aligned} \tag{6.4}$$

Wenn wir zudem annehmen, dass sich Kooperationsabbrüche unabhängig von der bisherigen Dauer der Transaktion ereignen, muss die Einschätzung $p_u(\gamma) = 1 - \exp(-\upsilon \cdot d(\gamma))$ lauten. Die Konstante υ gibt dabei ein, wie schnell mit steigender Transaktionsdauer die Wahrscheinlichkeit eines Kommunikationsabbruchs wächst. Eine Vereinfachung dieser Einschätzung ergibt sich dann, wenn die Ausführung aller im Informationssystem denkbaren Aktionen hinreichend gleich lange dauert. In diesem Fall unterscheidet sich die Dauer unterschiedlicher Transaktionen unwesentlich, so dass $p_u(\gamma)$ mit einer Konstante υ_0 angegeben werden kann.

Wie kommt der Systementwerfer zu einer Quantifizierung der Konstanten υ oder υ_0? Eine Möglichkeit besteht darin, dass er vor dem Einsatz des Informationssystems mit seinen eigenen Informationsgeräten experimentelle Erfahrungen über die Wahrscheinlichkeit von Kommunikationsabbrüchen macht. Die Auswertung dieser Erfahrungen ermöglicht es ihm, Aussagen über die Größe dieser Konstanten zu machen.

Abschließend stellt sich die Frage, wie außer den Kommunikationsabbrüchen andere Ursachen unbeabsichtigten Fehlverhaltens berücksichtigt werden können. Gemäß dem Systemmodell aus Abschnitt 1.2.2 fällt hierunter zum Beispiel das unvorhergesehene Ausschalten des Geräts, auf dem sich die Einheit befindet, etwa weil die Batterie entleert ist. Ein weiterer Faktor, der berücksichtigt werden könnte, stellen Veränderungen in der Umgebung dar. Bewegt sich zum Beispiel ein Gerät, so wird die Rate der Kommunikationsabbrüche sehr viel höher als im unbewegten Fall liegen. All diese Überlegungen lassen sich in die Einschätzung unbeabsichtigten Fehlverhaltens $p_u(\gamma)$ einbeziehen. Voraussetzung hierfür ist, dass die Wahrscheinlichkeit des Zutreffens dieser weiteren

Störfaktoren abgeschätzt werden kann. Beträgt zum Beispiel die Wahrscheinlichkeit für eine leere Batterie p_B, so erhalten wir eine neue Abschätzung $p'_u(\gamma) = p_B + \overline{p}_B \cdot p_u(\gamma)$. Dies ändert nichts am Konzept des TIB-Modells. Lediglich seine Genauigkeit wird dadurch verbessert. In der Evaluation des Kapitels 10 werden wir uns mit der Berücksichtigung von Kommunikationsabbrüchen begnügen, da diese im Umfeld der dort simulierten Ad-hoc-Netze eine herausragende Bedeutung besitzen.

6.3.3 Glaubensrevision

Bei der Besprechung des TIB-Modells sind wir im letzten Abschnitt von der Fragestellung ausgegangen, wie aus dem Typglauben ein Glaube über das Verhalten einer Einheit abgeleitet werden kann. Für die Glaubensrevision muss jedoch der umgekehrte Weg beschritten werden: Ausgehend von Transaktionserfahrungen, die sich auf das Verhalten einer Einheit beziehen, sind Rückschlüsse auf den Typ der Einheit zu ziehen.

In diesem Abschnitt beschäftigen wir uns mit einer solchen Glaubensrevision. Sie ergibt sich nicht nur aus der Verfügbarkeit neuer Transaktionserfahrungen sondern auch aus der Berücksichtigung von Typinformation und der Bewertung zuvor unbekannter Einheiten. Anstoß für die Revision des Typglaubens kann also von einem der drei folgenden Ereignisse gegeben werden:

- *Transaktionserfahrung:* In einer Transaktion wird kooperatives oder betrügendes Verhalten einer Einheit beobachtet.
- *Typinformation:* Über eine Einheit liegt die Information vor, von welchem Typ sie ist.
- *Neue Bekanntschaft:* Es wird die Bekanntschaft mit einer Einheit gemacht, deren Existenz bislang unbekannt war. In diesem Fall ist der initiale Typglaube über diese Einheit festzulegen.

Die Vorschriften der Glaubensrevision werden im Folgenden für diese drei Arten von Ereignissen ausgearbeitet. Dabei greifen wir auf die Odds-Darstellung aus Abschnitt 4.2 zurück, um die Revision des Typglaubens übersichtlich darzustellen. Die Besprechung geht wie bereits der letzte Abschnitt davon aus, dass sich Einheit X ihren Glauben über den Typ von Einheit Y bildet.

Berücksichtigung einer Transaktionserfahrung. Eine Transaktionserfahrung stellt eine Information über das Verhalten einer Einheit dar. Sie ist dadurch zu berücksichtigen, dass der Verhaltensglaube revidiert wird. Da dieser Verhaltensglaube im TIB-Modell aus dem Typglauben abgeleitet wird, folgern wir daraus, dass Transaktionserfahrungen zur Revision des Typglaubens führen müssen.

Bevor wir uns den eigentlichen Revisionsvorschriften zuwenden, müssen wir einen Zusammenhang zwischen der Art der Transaktionserfahrung und dem Verhalten des Transaktionspartners herstellen. Es gibt zwei Möglichkeiten für eine Transaktionserfahrung: Entweder wurde das Verhalten der Gegenübers Y als kooperativ ($C_Y^{(p)}$) oder als betrügend ($D_Y^{(p)}$) wahrgenommen (engl.: perception). Diese beiden Fälle stimmen nicht vollkommen mit dem eigentlichen Verhalten von Einheit Y ($C_Y^{(b)}$ und $D_Y^{(b)}$) überein. Der Unterschied liegt darin, dass sich Einheit X selber unbeabsichtigterweise fehlverhalten haben könnte. Gemäß dem Systemmodell aus Abschnitt 1.2.2 ist sie sich dessen nicht bewusst. Der Einheit Y erscheint dies jedoch als Betrug, so dass sie frühzeitig das Transaktionsprotokoll abbricht. Wenn aus der Sicht der Einheit X der nächste Schritt des Transaktionspartners Y ausbleibt, so könnte dies also auch an ihr selbst liegen. Die a priori Wahrscheinlichkeit dafür, dass eigenes unbeabsichtigtes Fehlverhalten eintritt, ist gemäß dem

TIB-Modell $p_u(\gamma)$. Der Zusammenhang zwischen Wahrnehmung des Verhaltens der Einheit Y und ihrem eigentlichen Verhalten ist somit wie folgt:

$$\begin{aligned} p_X(C_Y^{(p)},\gamma) &= p_X(C_Y^{(b)},\gamma)\cdot\overline{p}_u(\gamma) \\ p_X(D_Y^{(p)},\gamma) &= 1-p_X(C_Y^{(p)},\gamma) \end{aligned} \tag{6.5}$$

Die Revision des Typglaubens erfordert die Anwendung der Bayes-Formel. Dabei ist zu ermitteln, wie aus dem prioren Typglauben $p_X(N_Y)$ der posteriore Typglaube $p_X(N_Y|C_Y^{(p)},\gamma)$ und $p_X(N_Y|D_Y^{(p)},\gamma)$ für die Wahrnehmung von kooperativem beziehungsweise betrügendem Verhalten des Transaktionspartners im Kontext γ zu berechnen ist. Laut der Bayes-Formel benötigen wir hierfür die folgenden Wahrscheinlichkeiten, die sich aus den Formeln 6.1 und 6.2 des TIB-Modells und der Zusammenhangs mit der Wahrnehmung (Formel 6.5) ergeben:

$$\begin{aligned} p_X(C_Y^{(b)}|N_Y,\gamma) &= 1-p_u(\gamma)=\overline{p}_u(\gamma) \\ p_X(C_Y^{(p)}|N_Y,\gamma) &= p_X(C_Y^{(b)}|N_Y,\gamma)\cdot\overline{p}_u(\gamma)=\overline{p}_u(\gamma)^2 \\ p_X(D_Y^{(p)}|N_Y,\gamma) &= 1-\overline{p}_u(\gamma)^2 \\ p_X(C_Y^{(b)}|S_Y,\gamma) &= p_n(\gamma)\cdot(1-p_u(\gamma))=p_n(\gamma)\cdot\overline{p}_u(\gamma) \\ p_X(C_Y^{(p)}|S_Y,\gamma) &= \overline{p}_u(\gamma)\cdot p_X(C_Y^{(b)}|S_Y,\gamma)=p_n(\gamma)\cdot\overline{p}_u(\gamma)^2 \\ p_X(D_Y^{(p)}|S_Y,\gamma) &= 1-p_X(C_Y^{(p)}|S_Y,\gamma)=1-p_n(\gamma)\cdot\overline{p}_u(\gamma)^2 \end{aligned} \tag{6.6}$$

Die Revisionsvorschrift lässt sich übersichtlicher gestalten, wenn der Typglaube in der Odds-Darstellung $\hat{p}_X(N_Y)$ angegeben ist. In diesem Fall ist die Revision des Typglaubens durch die Multiplikation mit einem Revisionsfaktor r_C für kooperatives Verhalten und r_D für betrügendes Verhalten durchführen. Diese Faktoren berechnen sich wie folgt:

$$\begin{aligned} r_C(\gamma) &= \frac{p_X(C_Y^{(p)}|S_Y,\gamma)}{p_X(C_Y^{(p)}|N_Y,\gamma)}=\frac{p_n(\gamma)\cdot\overline{p}_u(\gamma)^2}{\overline{p}_u(\gamma)^2}=p_n(\gamma) \\ r_D(\gamma) &= \frac{p_X(D_Y^{(p)}|S_Y,\gamma)}{p_X(D_Y^{(p)}|N_Y,\gamma)}=\frac{1-p_n(\gamma)\cdot\overline{p}_u(\gamma)^2}{1-\overline{p}_u(\gamma)^2}=1+\frac{\overline{p}_n(\gamma)}{\frac{1}{\overline{p}_u(\gamma)^2}-1} \end{aligned} \tag{6.7}$$

Anhand dieser Faktoren lassen sich die Eigenschaften der Glaubensrevision kennzeichnen: **(1)** Wird kooperatives Verhalten wahrgenommen, so geht in die Glaubensrevision die Wahrscheinlichkeit für unbeabsichtigtes Fehlverhalten nicht ein. Dies ist probabilistisch korrekt, da bei Eintreten des Ereignisses $C_Y^{(p)}$ unbeabsichtigtes Fehlverhalten ausgeschlossen werden kann. **(2)** Wird betrügendes Verhalten wahrgenommen, so fällt die Revision umso schwächer aus, je größer die Wahrscheinlichkeit für unbeabsichtigtes Betrugsverhalten ist. Einen ähnlichen Sachverhalt haben wir bereits in Abschnitt 4.1.3 bei der Besprechung von *GTFT* gesehen. **(3)** Je wahrscheinlicher strategisches Einhalten angesehen wird, desto schwächer fällt die Revision aus. Dies ist auch bei der Wahrnehmung von Betrug sinnvoll, da dann die Wahrscheinlichkeit dafür höher ist, dass er unbeabsichtigt gewesen ist. **(4)** Die Revisionsfaktoren sind weder null noch unendlich, da gemäß dem Systemmodell weder beabsichtigtes noch unbeabsichtigtes Betrugsverhalten ausgeschlossen werden kann. Es gibt daher keine unmöglichen Ereignisse, die von den Revisionsvorschriften separat behandelt werden müssten.

Die Besprechung zeigt also, dass durch die vorgestellten Revisionsvorschriften Transaktionserfahrungen auf probabilistisch fundierte Weise berücksichtigt werden. Außerdem wird deutlich,

dass das dritte Entwurfsprinzip des Systementwurfs umgesetzt wird: Um als normativ angesehen zu werden, müssen strategische Einheiten von Betrugsverhalten absehen. Je unwahrscheinlicher unbeabsichtigtes Betrugsverhalten ist, desto schneller erscheint nämlich eine betrügende Einheit als strategisch.

Berücksichtigung von Typinformation. Typinformation lässt sich insofern einfacher als Verhaltensinformation berücksichtigen, als sie sich genauso wie der Glaube einer Einheit auf die Typ-Ebene bezieht. Dies schlägt sich in den Revisionsvorschriften nieder: Wird angezeigt, dass Einheit Y normativ beziehungsweise strategisch ist, so ergibt sich als posteriore Typglaube $p_X(N_Y|N_Y) = 1$ und $p_X(N_Y|S_Y) = 0$. Für die jeweiligen Revisionsfaktoren r_N und r_S gilt daher:

$$\begin{aligned} r_N &= \frac{p_X(S_Y|N_Y)}{p_X(N_Y|N_Y)} = \frac{0}{1} = 0 \\ r_S &= \frac{p_X(S_Y|S_Y)}{p_X(N_Y|S_Y)} = \frac{1}{0} = \infty \end{aligned} \tag{6.8}$$

Der Erhalt von Typinformation stellt nur dann ein unmögliches Ereignis dar, wenn Einheit X zuvor eine sichere Kenntnis von dem Typ der Einheit Y hatte und diese Kenntnis der erhaltenen Typinformation widerspricht. Dass dieser Fall nicht eintreten kann, ergibt sich daraus, dass Typinformationen die einzige Quelle für die sichere Kenntnis des Typs einer Einheit sind. Diese Typinformationen können sich jedoch nicht widersprechen, da sie nur dann berücksichtigt werden, wenn sie absolut zuverlässig sind. Diese Bedingung wird insbesondere von den Typbeweisen aus Abschnitt 7.3 erfüllt.

Berücksichtigung neuer Bekanntschaften. Wenn eine Einheit X die Bekanntschaft einer ihr zuvor unbekannten Einheit Y macht, so stellt sich die Frage, wie sie ihren Typglauben über diese Einheit Y initialisiert. Eine solche Initialisierung kann dabei nicht von Informationen über Einheit Y Gebrauch machen, da über bislang unbekannte Einheiten naturgemäß keine Informationen vorliegen. Als Anhaltspunkt für die Initialisierung steht immerhin die Tatsache zur Verfügung, dass diese Einheit Y am Informationssystem teilnimmt. Der Glaube über die Populationsstruktur des Gesamtsystems bestimmt daher, wie der Typglaube für eine zuvor unbekannte Einheit zu initialisieren ist. Ein solche *Systemglaube* ergibt aus der Kombination der zwei folgenden Quellen:

- *Interne Quellen:* Wenn Einheit X zuvor Erfahrungen mit einigen anderen Einheiten des Systems gemacht hat, so hat sie sich bereits einen Typglauben über diese Einheiten gebildet. Dieser Typglaube lässt sich zu einem gewissen Maße auf die Einheit Y übertragen. Wenn zum Beispiel Einheit X die meisten anderen Einheiten aufgrund ihrer Erfahrungen als strategisch einschätzt, so ist es wahrscheinlich, dass auch Einheit Y strategisch ist.

- *Externe Quellen:* Aufgrund des Umfeldes des Informationssystems ist der menschliche Prinzipal von Einheit X unter Umständen in der Lage, sich eine Vorstellung von dem zu erwartenden Anteil normativer Einheit zu bilden. Ein Anhaltspunkt hierfür bietet sich ihm darin, wie stark er die Manipulationskosten wahrnimmt und als wie manipulationsfreudig er die anderen Benutzer einschätzt. Im Campus-Szenario ist zum Beispiel denkbar, dass sich ein Student in seinem Bekanntenkreis zunächst über die Eigenheiten des Informationssystems umhört, bevor er sich zur Teilnahme am System entscheidet.

Offensichtlich sind externe Quellen für die Bildung des Systemglaubens weniger zuverlässig als interne Quellen. Auf sie kann dennoch nicht verzichtet werden, da einer Einheit am Anfang

ihrer Teilnahme am Informationssystem keinerlei Erfahrungen vorliegen und damit keine internen Quellen verfügbar sind.

Wie lassen sich diese Überlegungen in eine Quantifizierung des Systemglaubens umsetzen? Sei E_X die Menge von Einheiten, die Einheit X bereits kennt. Der Typglaube über diese Einheiten stellt die interne Quelle für den Systemglauben dar. Auf der anderen Seite sei $p_X(N_0)$ eine Einschätzung von Einheit X, wie hoch der Anteil der normativen Einheiten im System ist. Diese Einschätzung basiert auf den externen Quellen. Ein Ausgleich zwischen den internen und externen Quellen kann nur dann gefunden werden, wenn die einzelnen Quellen entsprechend ihrer Zuverlässigkeit gewichtet werden. Zu diesem Zweck kommt die Konstante α_0 zum Einsatz, die die Zuverlässigkeit der externen Quellen angibt. Die Quantifizierung der Konstante orientiert sich daran, wie aussagekräftig die Einschätzung $p_X(N_0)$ im Vergleich zu dem Typglauben über eine bereits bekannte Einheit ist. Konkret bedeutet dies für den Systemglauben $p_X(N_S)$ Folgendes:

$$p_X(N_S) = \frac{\alpha_0 \cdot p_X(N_0) + \sum_{e \in E_X} p_X(N_e)}{\alpha_0 + |E_X|} \tag{6.9}$$

Die Eigenschaften dieser Quantifizierung stellen sich also wie folgt dar: **(1)** Wenn die externen Quellen als absolut zuverlässig gelten, wird durch das Setzen $\alpha_0 = \infty$ der Typglaube über die bereits bekannten Einheiten ignoriert. In der Formel kommt dies dadurch zum Ausdruck, dass sich $p_X(N_S)$ zu $p_X(N_0)$ vereinfacht. Der Systemglaube ist also zu jedem Zeitpunkt gleich der externen Einschätzung über die relative Häufigkeit normativer Einheiten. **(2)** Wenn die externen Quellen vollkommen unzuverlässig sind, ist $\alpha_0 = \epsilon$ mit einem sehr kleinen Wert ϵ zu setzen. In diesem Fall wird nur bei der Bekanntschaft der ersten Einheit auf die externe Einschätzung $p_X(N_0)$ zurückgegriffen. Anschließend ergibt sich der Systemglaube aus dem Durchschnitt des Typglaubens über die bekannten Einheiten. Daraus folgt, dass der Systemglaube sich über die Zeit hinweg dynamisch den Erfahrungen der Einheit X anpasst. **(3)** Liegt der Wert α_0 zwischen diesen beiden Extremen, so gibt er die Zuverlässigkeit der externen Einschätzung als Äquivalenz davon an, über wie viele Einheiten ein Typglaube von $p_X(N_0)$ vorliegt. Ein Wert von $\alpha_0 = 2$ besagt zum Beispiel, dass bei der Bildung des Systemglaubens so getan wird, als ob es zwei "virtuelle" Einheiten gibt, über die der Typglaube genau die externe Einschätzung $p_X(N_0)$ beträgt.

Unter Zuhilfenahme des Systemglaubens lassen sich neue Bekanntschaften auf geradlinige Weise berücksichtigen: Lernt Einheit X die Einheit Y kennen, so setzt sie ihren initialen Typglauben $p_X(N_Y)$ über Einheit Y auf ihren aktuellen Systemglauben $p_X(N_S)$. Durch diese Initialisierung wird eine solide Basis für die weitere Glaubensbildung über den Typ von Einheit Y geschaffen.

6.4 Vertrauensentscheidungen

Der Zweck der Glaubensbildung besteht darin, Entscheidungen über die Teilnahme an Transaktionen als Vertrauensentscheidungen treffen zu können. In diesem Abschnitt beschäftigen wir uns damit, wie solche Vertrauensentscheidungen zu treffen sind. Hierfür entwickeln wir zunächst eine utilitaristische Bewertung der Rahmenbedingung einer Transaktion. Auf der Basis dieser Bewertung werden die Vorschriften dazu vorgestellt, wie Vertrauensentscheidungen getroffen werden.

Utilitaristische Bewertung von Rahmenbedingungen. Um eine Gelegenheit zu einer Transaktion bewerten zu können, muss der Nutzen ermittelt werden, der aus der Teilnahme an dieser Transaktion zu erwarten ist. Hierfür ist auf diejenigen Rahmenbedingungen der Transaktionsgelegenheit einzugehen, die einen Einfluss auf den erwarteten Nutzen ausüben. Bei diesen Aspekten der Rahmenbedingungen handelt es sich um die Folgenden:

- *Kosten-/Nutzenstruktur der Aktionen:* Eine Transaktion besteht darin, dass zwei Einheiten füreinander jeweils eine Aktion ausführen. Aus der Sicht einer dieser Einheiten sind daher zwei Größen zu berücksichtigen: Die Kosten für die Ausführung der eigenen Aktion a_o werden mit $c(a_o)$ notiert. Der Nutzen aus der Ausführung der Aktion a_p ihres Transaktionspartners bezeichnen wir mit $u(a_p)$.

- *Transaktionskosten:* Wenn bei einer Transaktion unabhängig von der Aktionsausführung Kosten anfallen, so bezeichnen wir sie mit c_T. Im Zwei-Wege Transaktionsprotokoll fallen diese Kosten dadurch an, dass jede Einheit je eine Nachricht senden und empfangen muss. Im Sechs-Wege Transaktionsprotokoll aus dem nachfolgenden Kapitel 7 kommen noch das Versenden und Empfangen von zwei weiteren Nachrichten und vier kryptographische Operationen hinzu.

- *Position:* Es macht für eine Einheit einen Unterschied, ob sie in einer Transaktion die Risiko-Position oder die sichere Position einnimmt.

- *Transaktionskontext:* Zur Einschätzung des wahrscheinlichen Verhaltens des Transaktionspartners wird die Angabe des Transaktionskontextes γ benötigt.

Wie ist aufgrund dieser Größen der erwartete Nutzen einer Transaktionsgelegenheit zu ermitteln? Hierbei müssen wir zwischen den verschiedenen Positionen einer Einheit unterscheiden. Befindet sich die Einheit in der Risiko-Position, so muss sie zuerst ihre Aktion ausführen. Daher fallen für sie die Kosten für die Aktionsausführung auf jeden Fall an. Allerdings erhält sie den Nutzen aus der Aktionsausführung des Transaktionspartners nur, wenn sein Verhalten als kooperativ wahrgenommen wird[2]. Wenn wir zudem die Transaktionskosten als Fixkosten ansehen, so erhalten wir für den erwarteten Nutzen u_R in der Risiko-Position folgende Einschätzung (aus der Sicht von Einheit X mit dem potentiellen Transaktionspartner Y):

$$u_R(a_o, a_p, \gamma) = p_X(C_Y^{(p)}|\gamma) \cdot u(a_p) - c(a_o) - c_T \tag{6.10}$$

Befindet sich eine Einheit in der sicheren Position, so führt sie ihre Aktion nur dann aus, wenn der Transaktionspartner zuvor die seinige ausgeführt hat. Daher erhalten wir für den erwarteten Nutzen u_S in der sicheren Position:

$$u_S(a_o, a_p, \gamma) = p_X(C_Y^{(p)}|\gamma) \cdot [u(a_p) - c(a_o)] - c_T \tag{6.11}$$

Der Unterschied zwischen diesen Einschätzung besteht also einzig darin, dass die Kosten für die eigene Aktionsausführung nur unter gewissen Umständen anfallen. Damit bestätigen die Formeln des erwarteten Nutzens unsere Intuition, dass das Innehaben der sicheren Position Vorteile bietet. Aufgrund dessen ist zu erwarten, dass die Einheiten bei der Aushandlung der Rahmenbedingung der Transaktion auf die Zuweisung der Position einen ebenso großen Wert legen, wie auf die auszuführenden Aktionen.

In einem Punkt müssen wir von der utilitaristischen Bewertung von Transaktionen abweichen: Das zweite Entwurfsprinzip des Systementwurfs fordert, dass strategische Einheiten als normativ erscheinen wollen. Die einzige Möglichkeit hierfür besteht darin, dass nur mit normativ

[2]Kooperatives Verhalten durch den Transaktionspartner alleine reicht nicht aus, um in den Genuss seiner Aktionsausführung zu kommen. Es könnte nämlich auch sein, dass die Einheit sich selbst unbeabsichtigterweise fehlverhalten hat. In diesem Fall wird auch ein kooperativer Transaktionspartner seine Aktion nicht ausführen. Daher richten wir die Vertrauensentscheidungen an der Wahrscheinlichkeit aus, dass kooperatives Verhalten auch tatsächlich wahrgenommen wird. Dadurch lässt sich der Fall eigenen unbeabsichtigten Fehlverhaltens ausschließen.

erscheinenden Einheiten Transaktionen eingegangen werden. Hingegen sieht die Berechnung von $p_X(C_Y^{(p)})$ vor, dass selbst Einheiten, die mit Sicherheit strategisch sind, sich unter Umständen kooperativ verhalten. Als Folge davon werden solche strategischen Einheiten nicht von weiteren Transaktionen ausgeschlossen. Eben dies ist aber notwendig, um die Betrugskosten hochzuhalten und das zweite Entwurfsprinzip umzusetzen. Daraus ergibt sich, dass die Ermittlung des erwarteten Nutzens einer Transaktion angepasst werden muss. Dabei ist die Wahrscheinlichkeit kooperativen Verhaltens des Transaktionspartners daran auszurichten, wie wahrscheinlich er normativ ist und als solcher als kooperativ wahrgenommen wird. In der Formel ist also $p_X(C_Y^{(p)})$ mit $p_X(N_Y) \cdot p_X(C_Y^{(p)}|N_Y)$ zu ersetzen. Wir erhalten demgemäß die folgenden Formeln:

$$\begin{aligned} u_R(a_o, a_p, \gamma) &= p_X(N_Y) \cdot \overline{p}_u(\gamma)^2 \cdot u(a_p) - c(a_o) - c_T \\ u_S(a_o, a_p, \gamma) &= p_X(N_Y) \cdot \overline{p}_u(\gamma)^2 \cdot [u(a_p) - c(a_o)] - c_T \end{aligned} \tag{6.12}$$

Treffen von Vertrauensentscheidungen. Die Bewertung der Rahmenbedingungen einer Transaktionsgelegenheit bildet die Grundlage für das Treffen von Vertrauensentscheidungen. Für die Entscheidung über die Teilnahme an einer Transaktion bedeutet dies konkret: Zunächst sind all jene Transaktionsgelegenheiten abzulehnen, bei denen der erwartete Nutzen negativ ist. Dieses Kriterium alleine reicht jedoch nicht aus. Dies liegt daran, dass es mehrere Transaktionsgelegenheiten geben kann, die sich gegenseitig ausschließen. Diese Situation tritt zum Beispiel im Campus-Szenario auf, wenn Claudes Einheit am Mensaplan interessiert ist und nicht nur Davids Einheit einen solchen anbietet. In diesem Fall muss Claudes Einheit wählen, mit welcher anderen Einheit eine Transaktion zum Erhalt des Mensaplans eingegangen wird. Diese Wahl richtet sich nach einem zweiten Kriterium: Gibt es mehrere Transaktionsgelegenheiten, die sich gegenseitig ausschließen, so wird diejenige gewählt, deren erwarteter Nutzen am höchsten ist.

Aus der Sicht des *IPD/CR*-Spiels aus Abschnitt 4.1.3 definieren die Vertrauensentscheidungen das Wahlverhalten im Vorfeld des Gefangenendilemmas. Damit ist die Strategie der normativen Einheiten im *IPD/CR*-Spiel wie folgt: Die Wahl der Transaktionspartner erfolgt in Abhängigkeit des eigenen Glaubens gemäß dem jeweiligen erwarteten Nutzen. Abgelehnt werden von Vornherein diejenigen potentiellen Transaktionspartner, bei denen ein negativer Nutzen erwartet wird. Erst diese Möglichkeit zur Ablehnung erlaubt es den normativen Einheiten, in ihren Transaktionen gemäß der ersten Norm nie Betrugsverhalten zu beabsichtigen, ohne dass sie sich allzu altruistisch verhalten würden. Damit wird auch die zweite Maxime des Systementwurfs eingehalten, dass die Nachteile von normativem Verhalten gering gehalten werden.

6.5 Zusammenfassung

Als Ausgangspunkt für die Vertrauensbildung macht jede Einheit durch Teilnahme an Transaktionen ihre eigenen Erfahrungen über das Verhalten Anderer. Dieses Kapitel hat den Entwurf einer solchen lokalen Vertrauensbildung vorgestellt. Hierfür wurde zunächst ihr Kreislauf um einige Erweiterungen ergänzt und eine datenzentrische Sicht auf ihn gegeben. Der erste Schritt des Kreislaufs besteht in der Teilnahme an *Transaktionen*. Den zeitlichen Ablauf einer Transaktion haben wir im Zwei-Wege Transaktionsprotokoll festgelegt. Zudem wurde die wichtigste Norm des Systementwurfs vorgestellt, die die erste Maxime des Systementwurfs umsetzt. Sie orientiert sich am Transaktionsverhalten der Einheiten und besagt, dass Einheiten ihre Transaktionspartner nicht betrügen sollen.

Den Schwerpunkt dieses Kapitels bildete die *Glaubensbildung*. Es wurde gezeigt, dass sie typorientiert sein muss, um die drei Anforderungen an sie bewältigen zu können. Darin unterscheidet sie sich grundsätzlich von der verhaltensorientierten Glaubensbildung der verwandten Arbeiten. Diese Überlegungen wurden in einem Glaubensmodell namens TIB-Modell zusammengefasst. Es stellt eine Beziehung zwischen dem Typ und dem Verhalten einer Einheit indirekt über ihre Absicht her. Diese Indirektion wird aufgrund der Möglichkeit strategischer Einhaltung von Normen und unbeabsichtigten Betrugsverhaltens notwendig. Zur Einschätzung der Wahrscheinlichkeit strategischer Einhaltung wurde eine Formel abgeleitet, die sich auf dem Transaktionswert als den einzigen beobachtbaren Faktor strategischen Verhaltens abstützt. Weiterhin wurde unbeabsichtigtes Betrugsverhalten gemäß der Wahrscheinlichkeit für Kommunikationsabbrüche eingeschätzt. Die Vorschriften der Glaubensrevision leiten sich aus dem TIB-Modell ab. Ausgehend von Transaktionserfahrungen, die sich auf das Verhalten einer Einheit beziehen, muss sie Rückschlüsse auf den Typ der Einheit ziehen. Hierfür wurden entsprechende Revisionsfaktoren angegeben, deren Anwendung zu einer probabilistisch fundierten Glaubensbildung führen. Damit wurde das dritte Entwurfsprinzip des Systementwurfs dadurch umgesetzt, dass betrügende Einheiten aufgrund dieser Glaubensrevision strategischer erscheinen. Außerdem wurden Revisionsvorschriften für die Berücksichtigung von Typinformationen hergeleitet. Zur Bewertung neuer Bekanntschaften wird der Typglauben entsprechend des Systemglaubens initialisiert. Zu diesem Zweck haben wir besprochen, wie der Systemglauben zu quantifizieren ist.

Zum Treffen von *Vertrauensentscheidungen* wurden Formeln abgeleitet, mit deren Hilfe der erwartete Nutzen einer Transaktionsgelegenheit bewertet werden kann. Diese Bewertung berücksichtigt das zweite Entwurfsprinzip des Systementwurfs, indem sie strategische Einheiten schlechter stellt. Die Entscheidung über die Teilnahme an einer Transaktion wird aus dieser Bewertung abgeleitet. Erst die Möglichkeit zur Ablehnung von potentiellen Transaktionspartnern erlaubt es den normativen Einheiten, sich trotz des Verzichts auf Betrugsabsichten nicht altruistisch zu verhalten. Damit wurde auch die zweite Maxime des Systementwurfs eingehalten.

Kapitel 7

浸潤之譖 膚受之訴 不行焉 可謂明也已矣

"Derjenige, der sich weder zum Durchsickern von Verleumdungen noch zu tätlichen Beschuldigungen verleiten lässt, kann in der Tat der Einsichtige genannt werden."
(Gespräche und Aussprüche des Konfuzius, 12.6)

Erweiterung um transaktionale Beweismittel

Die lokale Vertrauensbildung geht davon aus, dass jede Einheit durch Teilnahme an Transaktionen ihre eigenen Erfahrungen über das Verhalten Anderer macht. Basierend auf diesen Erfahrungen bildet sich jede Einheit ihren Glauben und fällt ihre Vertrauensentscheidungen. Diese Glaubensbildung kann auf eine breitere Grundlage gestellt werden, wenn einer Einheit auch die Transaktionserfahrungen Anderer zugänglich gemacht werden. Zu diesem Zweck müssen die Einheiten sich untereinander über ihre Erfahrungen berichten. Dadurch verstärkt sich die gegenseitige Kontrolle der autonomen Einheiten. Die sich ergebende Vertrauensbildung ist insofern verteilt, als in sie nicht nur die eigenen Transaktionserfahrungen sondern auch die der anderen Einheiten einfließen. Als Voraussetzung hierfür ist allerdings sicherzustellen, dass die Berichte über Transaktionserfahrungen wahrheitsgemäß und verfügbar sind.

Mit dieser Aufgabe beschäftigen wir uns in diesem Kapitel. Zu diesem Zweck führen wir in Abschnitt 7.1 Beweismittel als das entscheidende Hilfsmittel ein. In Abschnitt 7.2 identifizieren wir zwei Arten von transaktionalen Beweismitteln und legen dar, wie sie in ein erweitertes Transaktionsprotokoll eingehen. Dies ermöglicht in Abschnitt 7.3 die Konzeption eines Empfehlungssystems, das auf die transaktionalen Beweismitteln ausgerichtet ist. Die Empfehlungen beinhalten Berichte über Transaktionserfahrungen und unterstützen die Einheiten dabei, ihren Glauben über den Typ Anderer zu bilden. Die entsprechenden Vorschriften der Glaubensrevision werden in Abschnitt 7.4 vorgestellt. Als Bindeglied zwischen dem Transaktionsprotokoll, dem Empfehlungssystem und der Glaubensbildung agiert die Komponente zur Verwaltung von Beweismitteln und Wissen, auf die wir in Abschnitt 7.5 eingehen. Abschließend analysieren wir in Abschnitt 7.6, ob und wie das Verhalten strategischer Einheiten von den Vorschriften des Entwurfs abweicht.

7.1 Einführung

Im Folgenden geben wir eine Übersicht der Funktionsweise der verteilten Vertrauensbildung. Hierfür gehen wir zunächst in Abschnitt 7.1.1 auf das Konzept der Beweismittel ein, das die Vertrauensbildung auf entscheidende Weise unterstützt. Anschließend stellen wir in Abschnitt 7.1.2 dar, wie der Kreislauf der Vertrauensbildung zu erweitern ist, wenn transaktionale Beweismittel zum Einsatz kommen.

7.1.1 Konzept der Beweismittel

In diesem Abschnitt stellen wir das Konzept der Beweismittel vor. Zu diesem Zweck definieren wir, was unter Beweismitteln zu verstehen ist, und stellen heraus, welche Eigenschaften der Beweismittel unmittelbar aus der Definition folgen. Anschließend zeigen wir, dass Beweismittel Möglichkeiten zur glaubwürdigen Signalisierung schaffen. Daraus erklärt sich, warum der Kreislauf der Vertrauensbildung vom Einsatz von Beweismitteln profitiert und daher entsprechend erweitert werden muss. Auf eine solche invasive Anwendung der Beweismittel gehen wir zum Abschluss dieses Abschnitts ein.

Definition und unmittelbare Eigenschaften. In Abschnitt 2.1 haben wir bereits die Techniken der Nichtabstreitbarkeit kennen gelernt. Sie erlauben es, dass jede Einheit zum Signieren von Informationen in der Lage ist und diese Signatur von jeder anderen Einheit auf ihre Richtigkeit hin überprüft werden kann. Eine signierte Information stellt insofern eine *nicht-abstreitbare Marke* (engl.: non-repudiable token) dar, als die signierende Einheit anderen Einheiten gegenüber nicht abstreiten kann, dass sie diese Information signiert hat. Kommt es dennoch zu einem solchen Abstreiten, so genügt ein Vorzeigen der Signatur dazu, die anderen Einheiten davon zu überzeugen, dass die abstreitende Einheit lügt. Im Folgenden nennen wir diejenige Einheit, die eine Information signiert, den *Aussteller* der sich dadurch ergebenden nicht-abstreitbaren Marke.

Ein *Beweismittel* ist nichts anderes als eine nicht-abstreitbare Marke. Die Information, die in der nicht-abstreitbaren Marke signiert wird, nennen wir die *Aussage* des Beweismittels. Durch die Ausstellung eines Beweismittels legt sich eine Einheit also auf eine Aussage fest. Diese *Festlegung* (engl.: commitment) besitzt drei entscheidende Eigenschaften:

- *Übertragbarkeit:* Jede Einheit kann von dieser Festlegung erfahren. Hierfür muss es lediglich zu einer Übertragung des Beweismittels kommen. Jedes Beweismittel ist über den Kommunikationskanal übertragbar, da sowohl die Information als auch die Signatur einer jeden nicht-abstreitbaren Marke in elektronischer Form vorliegt.

- *Überprüfbarkeit:* Jede Einheit ist in der Lage, bei Erhalt eines Beweismittels zu erkennen, wer das Beweismittel ausgestellt hat und zu welcher Aussage er sich festgelegt hat. Dies wird durch die kryptographischen Eigenschaften der Signatur zugesichert. Insbesondere kann ausgeschlossen werden, dass eine Einheit unter dem Namen einer anderen Einheit ein Beweismittel ausstellt.

- *Integrität:* Außer dem Aussteller selbst ist keine andere Einheit in der Lage, die Aussage eines Beweismittels zu verändern, ohne dass andere Einheiten dies bemerken. Erhält eine Einheit ein Beweismittel von einer anderen Einheit, so kann sie das Beweismittel zwar bei sich lokal ablegen oder anderen Einheiten weitergeben. Eine Manipulation des Beweismittels ist in beiden Fällen jedoch ausgeschlossen.

Durch diese drei Eigenschaften ergeben sich zwei bemerkenswerte Folgen: Zum einen kann die Festlegung einer Einheit jeder beliebigen anderen Einheit *glaubhaft* vermittelt werden. Dazu ist lediglich das Vorzeigen des entsprechenden Beweismittels notwendig. Gerade in Informationssystemen mit beschränkten Kommunikationsmöglichkeiten wie im Ad-hoc Netz des Campus-Szenarios ist diese Eigenschaft von Vorteil: Die Aussage einer Einheit kann nämlich auch in ihrer Abwesenheit nach Bedarf glaubhaft reproduziert werden. Zum anderen stellt die Manipulationssicherheit der Beweismittel eine passende Ergänzung für den Systementwurf dar, da dieser unter der Möglichkeit der Manipulation der originalen Systemsoftware leidet.

Auf den ersten Blick erscheint es fraglich, warum nur Beweismittel aber nicht die Systemsoftware manipulationssicher gemacht werden können. Dieser Widerspruch löst sich jedoch auf, wenn wir uns fragen, von wem jeweils eine solche Manipulation durchgeführt wird:

- *Jeweiliger menschlicher Prinzipal:* Bei der Systemsoftware manipuliert der menschliche Prinzipal einer Einheit. Er ist in der Lage, dies zu tun, da sein Informationsgerät, auf dem sich seine Einheit befindet, unter seiner Kontrolle ist. Bei der Systemsoftware ist also die Möglichkeit zur Manipulation nur deswegen gegeben, weil jede Einheit ihrem jeweiligen menschlichen Prinzipal gegenüber nicht autonom ist.

- *Einheiten untereinander:* Bei Beweismitteln bestünde hingegen eine Manipulation darin, dass eine Einheit A das von einer anderen Einheit B ausgestellte Beweismittel manipuliert. Da die Beweismittel durch die Techniken der PKI (vergleiche Abschnitt 2.1) kryptographisch gesichert sind, ist hierfür der private Schlüssel erforderlich, den jedoch nur Einheit B besitzt. Diese Einheit B ist in der Lage, der Einheit A den Zugriff auf den Schlüssel zu verwehren, weil sie ihr gegenüber autonom ist. Diese Überlegung zeigt, dass die Manipulationssicherheit von Beweismitteln direkt von den Zusicherungen der kryptographischen Verfahren herrührt. Solange diese Verfahren nicht scheitern, können Einheiten ihre Beweismittel nicht gegenseitig manipulieren.

Durch den Einsatz von Beweismitteln verändert sich das Systemmodell wie folgt: Jede Einheit ist in der Lage, sich auf bestimmte Aussagen festzulegen. Diese Festlegung ist jeder anderen Einheit glaubwürdig vermittelbar, indem ihr das Beweismittel zugesandt wird, das diese Festlegung enthält.

Auf der anderen Seite gibt es eine Reihe von an sich wünschenswerten Eigenschaften, die durch den Einsatz von Beweismitteln nicht zugesichert werden können. Bei ihnen handelt es sich um folgende Punkte:

- *Mehrseitige Festlegungen:* Die Ausstellung eines Beweismittels erfolgt durch genau eine Einheit. Daher können Festlegungen immer nur einseitig sein. Wenn sich mehrere Einheiten zu einer Aussage festlegen wollen, dann kann dies also nur durch eine Reihe einseitiger Festlegungen dieser Einheiten geschehen. Das Problem des koordinierten Angriffs aus Abschnitt 4.2 zeigt, dass unabhängig des verwendeten Protokolls die beteiligten Einheiten kein gemeinsames Wissen darüber erlangen können, ob es zur mehrseitigen Festlegung gekommen ist. Dies werden wir beim Entwurf des Quittungsmechanismus in Abschnitt 7.2 berücksichtigen.

- *Verbreitung negativer Aussagen:* Unter einer negativen Aussage verstehen wir eine Aussage, die einer anderen Einheit betrügendes Verhalten bescheinigt. Trägt ein Beweismittel eine solchen negative Aussage, so können wir nicht erwarten, dass es von der Einheit, auf die sich die Aussage bezieht, an andere Einheiten weiter gegeben wird. Um ein solches Beweismittel

im System zu verbreiten, muss es also von seinem Aussteller an unbeteiligte Einheiten weiter gegeben werden. Dies wird beim Entwurf von negativen Empfehlungen in Abschnitt 7.3 berücksichtigt. Bei positiven Aussagen ergibt sich dieses Problem so nicht, da die Einheit, auf die sich diese Aussage bezieht, zur Verbreitung des Beweismittels herangezogen werden kann.

- *Wahrheitsgehalt von Aussagen:* Der Einsatz von Beweismitteln ermöglicht lediglich, Festlegungen glaubhaft zu vermitteln. Die Aussage der Festlegung unterliegt dabei keiner Einschränkung. Eine Einheit ist also jederzeit dazu in der Lage, sich auf eine Aussage festzulegen, die ihrem Wissensstand widerspricht. Wir berücksichtigen dies bei den Revisionsvorschriften für die Bewertung von Selbstempfehlungen in Abschnitt 7.4.

Der letzte Punkt wirft die Frage auf, welcher Nutzen sich aus dem Einsatz von Beweismitteln ergibt. Die Einheiten sind nämlich nach wie vor in der Lage, Aussagen zu treffen, die nicht wahrheitsgemäß sind. Der folgende Paragraph klärt diese Frage, indem er zeigt, dass Beweismittel zwecks der Schaffung von Signalisierungsmöglichkeiten eingesetzt werden.

Signalisierung durch Beweismittel. Stellt eine Einheit ein Beweismittel aus, so kann jede andere Einheit nachvollziehen, zu welcher Aussage die Einheit sich festgelegt hat. Damit signalisiert der Aussteller eines Beweismittels, dass er bereit ist, für seine Aussage einzustehen. Je nach dem, ob sich eine Aussage im Nachhinein bewahrheitet, wird die Einheit, die sich zur Aussage festgelegt hat, von den anderen Einheiten bewertet. In Beweismitteln getroffene Aussagen stellen damit keinesfalls *billiges Gerede* im Sinne von Abschnitt 4.1.2 dar. Dies lässt sich an den beiden Bedingungen der glaubhaften Signalisierung festmachen:

- *Kann-Bedingung:* Eine Einheit, die eine wahrheitsgemäße Aussage treffen will, kann dies tun, indem sie ein entsprechendes Beweismittel ausstellt.

- *Kann-Nicht-Bedingung:* Legt sich eine Einheit zu einer Aussage fest, die sich nicht bewahrheitet, so wird sie in den Augen anderer Einheiten abgewertet und seltener als Transaktionspartner akzeptiert. Die etwaigen Vorteile von Falschaussagen werden dadurch unter Umständen zunichte gemacht. Daher erscheint eine Signalstörung einer Einheit nicht vorteilhaft.

Durch den Einsatz von Beweismitteln werden also Möglichkeiten zur glaubwürdigen Signalisierung geschaffen. Abbildung 7.1 stellt schematisch dar, auf welche Arten diese Möglichkeiten genutzt werden können. Bei ihnen handelt es sich um die folgenden zwei Arten:

- *Direkte Signalisierung:* Einheit A trifft eine Aussage gegenüber Einheit B. Um ihrer Aussage Nachdruck zu verschaffen, legt sich Einheit A auf diese Aussage fest, indem sie ein entsprechendes Beweismittel ausstellt und der Einheit B übergibt. Dadurch signalisiert sie Einheit B gegenüber, dass sie vom Zutreffen der Aussage überzeugt ist.

- *Indirekte Signalisierung:* Wenn Einheit A Einheit B von einem Sachverhalt überzeugen möchte, kann sie hierfür die Aussage einer dritten Einheit C anführen. Zu diesem Zweck muss Einheit A nicht nur eine entsprechende Aussage der Einheit C sondern auch ihre Festlegung hierzu in Form eines Beweismittels vorliegen. Mit Hilfe dieses Beweismittels kann Einheit A Einheit B gegenüber signalisieren, dass der Sachverhalt zutrifft. Diese Signalisierung ist insofern indirekt, als sie auf einem Beweismittel basiert, das nicht von der signalisierenden Einheit selbst ausgestellt wurde.

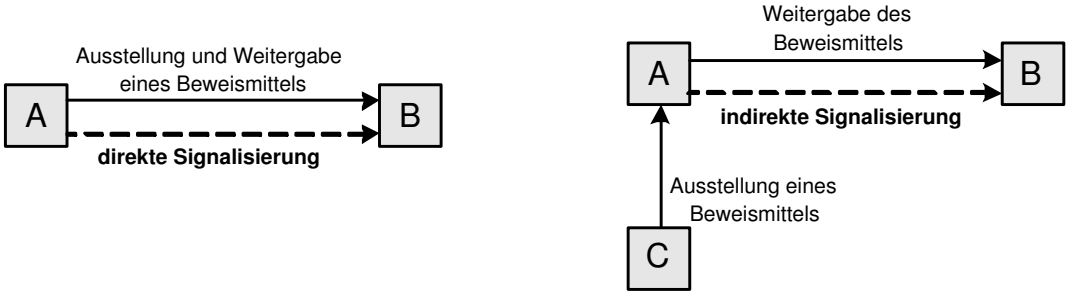

Abbildung 7.1: Direkte und indirekte Signalisierung mit Beweismitteln

Beide Arten der Signalisierung komplettieren sich gegenseitig. Wenn eine Einheit durch ihre eigene Festlegung die andere Einheit überzeugen kann, so reicht die direkte Signalisierung aus. Andernfalls ist die indirekte Signalisierung anzuwenden.

Invasive Anwendung auf den Kreislauf der Vertrauensbildung. Die bisherige Besprechung von Beweismitteln ist weitgehend abstrakt geblieben. Im Folgenden untersuchen wir, wie der Kreislauf der Vertrauensbildung vom Einsatz der Beweismittel profitieren kann.

Im Rahmen der Vertrauensbildung sind zwei Arten von Aussagen von Interesse, nämlich solche über den *Typ* und solche über das *Transaktionsverhalten* einer Einheit. Mit Hilfe dieser Aussagen werden die Einheiten in ihrer Glaubensbildung unterstützt, da ihre jeweiligen Transaktionserfahrungen dann nicht mehr ihre einzige Informationsquelle darstellen. Werden diese Aussagen im Zuge der Ausstellung von Beweismitteln getroffen, so sind Einheiten in der Lage, ihre Sicht über den Typ und das Transaktionsverhalten anderer Einheiten zu signalisieren.

Es ergeben sich daher zwei prinzipielle Ausprägungen von Beweismitteln: Die Aussage eines *transaktionalen Beweismittels* bezieht sich auf das Transaktionsverhalten einer Einheit. Mit solchen Beweismitteln beschäftigen wir uns in diesem Kapitel 7. Hingegen werden Aussagen über den Typ einer Einheit in *sozialen Beweismitteln* getroffen. Diese Beweismittel werden im darauffolgenden Kapitel 8 behandelt. Die Motivation hinter der Bezeichnungsweise der sozialen Beweismittel wird dabei deutlich werden.

Kommen Beweismittel zum Einsatz, so muss der Kreislauf der Vertrauensbildung erweitert werden. Diese invasive Anwendung von Beweismitteln erfolgt entlang der folgenden Schritte:

1. *Ausstellung von Beweismitteln:* Zunächst sind Arten von Beweismitteln zu finden, deren Aussagen für andere Einheiten von Interesse ist. Wir werden zum Beispiel in diesem Kapitel drei verschiedene Arten von transaktionalen Beweismitteln identifizieren.

2. *Einbeziehung in das Empfehlungssystem:* Das Kernproblem existierender Ansätze zur verteilten Vertrauensbildung liegt darin, dass ein Empfehler freie Wahl über den Inhalt der auszustellenden Empfehlung hat. Dieses Problem kann nur dann gelöst werden, wenn wir das Empfehlen als eine Möglichkeit zum indirekten Signalisieren ansehen. Damit eine Einheit empfehlen kann, müssen ihr also entsprechende Festlegungen anderer Einheiten in Form von Beweismitteln vorliegen. Daher bestimmt die Art der eingesetzten Beweismittel, welche prinzipiellen Empfehlungsmöglichkeiten gegeben sind. Als Folge davon ist ein Empfehlungssystem zu entwerfen, welches an den verwendeten Beweismitteln ausgerichtet ist.

3. *Einbeziehung in die Glaubensrevision:* Beweismittel und Empfehlungen zielen letztendlich darauf, dass die Erfahrungen und Sichtweisen einer Einheit in die Glaubensbildung anderer

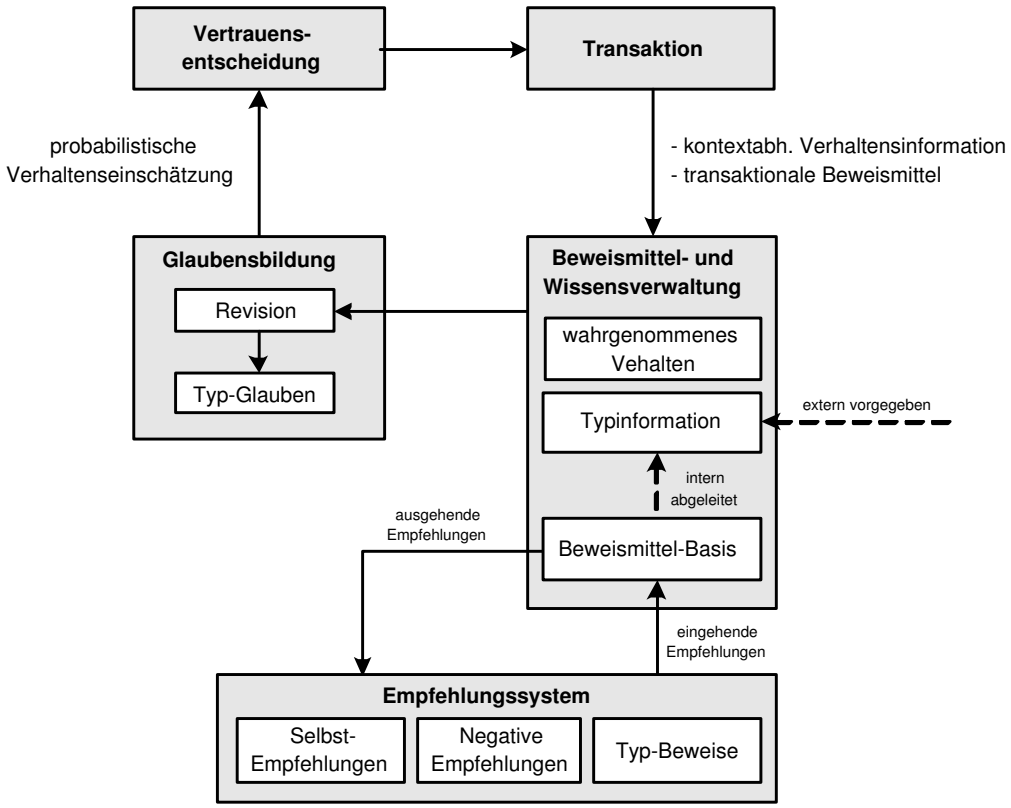

Abbildung 7.2: Kreislauf der verteilten Vertrauensbildung mit transaktionalen Beweismitteln

Einheiten mit einfließen. Dazu ist festzulegen, wie der Typglauben einer Einheit bei Erhalt einer Empfehlung zu revidieren ist.

7.1.2 Transaktionale Beweismittel im Kreislauf der Vertrauensbildung

Im Folgenden geben wir eine Übersicht, wie sich die invasive Anwendung von transaktionalen Beweismitteln auf den Kreislauf der Vertrauensbildung auswirkt. Dabei zeigen wir analog zu Abschnitt 6.1 zwei verschiedene Sichten auf den Kreislauf, nämlich eine funktionszentrische und eine datenzentrische Sicht.

Funktionszentrische Sicht. Ausgangspunkt für eine Erweiterung um transaktionale Beweismittel ist der Kreislauf der lokalen Vertrauensbildung. Bereits bei diesem Kreislauf ergaben sich laut Kapitel 6 einige Unterschiede zu den verwandten Arbeiten. Darüber hinaus führt der Einsatz von Beweismitteln und der daraus folgende Entwurf des Empfehlungssystems zu einem Kreislauf der verteilten Vertrauensbildung, der sich grundsätzlich von der Vertrauensbildung verwandter Arbeiten unterscheidet.

Der Kreislauf der verteilten Vertrauensbildung, der sich durch die Erweiterung um transaktionale Beweismittel ergibt, wird in Abbildung 7.2 gezeigt. Er unterscheidet sich vom Kreislauf der lokalen Vertrauensbildung in den folgenden Punkten:

- *Austausch von transaktionalen Beweismitteln:* Im Zuge einer Transaktion führen Einheiten nicht nur füreinander Aktionen aus. Hinzu kommt das gegenseitige Ausstellen von transak-

tionalen Beweismitteln. Als Ergebnis einer Transaktion erhält eine Einheit damit nicht nur eine Information über das Verhalten ihres Transaktionspartners. Darüber hinaus stehen ihr auch die Beweismittel zur Verfügung, die der Transaktionspartner während der Transaktion ausgestellt hat.

- *Erweiterung der Wissensverwaltung:* Die verfügbaren Beweismittel werden in einer Beweismittel-Basis abgelegt. Als Folge davon ist die ursprüngliche Wissensverwaltung zu einer kombinierten Beweismittel- und Wissensverwaltung auszubauen. Im Gegensatz zum Kreislauf der lokalen Vertrauensbildung erlangt sie große Bedeutung, da Schlüsse auf der Basis von Beweismitteln und Wissen gezogen werden können. Diese Schlussfolgerungen stehen als zusätzliche Eingabe der Glaubensrevision zur Verfügung. Eine besondere Schlussfolgerung stellt ein Beweis über den Typ einer Einheit dar. Er ist eine interne Quelle für Typinformationen.

- *Beweismittel-basiertes Empfehlungssystem:* Die Beweismittel-Basis bildet die Grundlage für den Umgang mit Empfehlungen. Da jede Empfehlung eine indirekte Signalisierung darstellt, wird ein entsprechendes Beweismittel benötigt, um sie zu unterstützen. Ein Kennzeichen des Empfehlungssystems ist also, dass Empfehlungen im Gegensatz zu den verwandten Arbeiten keine Glaubensberichte sind sondern Beweismittel enthalten. Das Ausstellen von Empfehlungen erfordert damit den Zugriff auf die Beweismittel-Basis. Auf der anderen Seite werden die Beweismittel, die in eingehenden Empfehlungen enthalten sind, in der Beweismittel-Basis abgelegt. Aufgrund der verschiedenen zum Einsatz kommenden transaktionalen Beweismittel unterscheiden wir drei Arten von Empfehlungen, nämlich Selbstempfehlungen, negative Empfehlungen und Typbeweise.

- *Beweismittel-basierte Glaubensrevision:* Die zusätzlichen Informationsquellen, die durch den Einsatz von Beweismitteln erschlossen werden, können nur dann Eingang in die Glaubensbildung finden, wenn entsprechende Revisionsvorschriften ausgearbeitet werden.

Diese Erweiterung des Kreislaufs der Vertrauensbildung ermöglicht es, dass die drei Anforderungen an das Empfehlungssystem aus Abschnitt 2.4.2 erfüllt werden können. Der *Wahrheitsbezug* der Empfehlungen leitet sich direkt aus der Glaubwürdigkeit der Beweismittel-basierten Signalisierung ab. Dadurch sind wir in der Lage, eine *fundierte Glaubensrevision* für die Einschätzung von Empfehlungen vorzuschlagen. Durch eine geeignete Auslegung der Revisionsvorschriften schaffen wir einen Anreiz für das Ausstellen von Empfehlungen. Damit wird auch die Forderung nach der *Verfügbarkeit* der Empfehlungen erfüllt.

Datenzentrische Sicht. Die Auswirkungen, die durch den Einsatz transaktionaler Beweismittel entstehen, werden besonders deutlich, wenn wir die datenzentrische Sicht des Kreislaufs der Vertrauensbildung einnehmen.

In Abbildung 7.3 wird diese Sicht dargestellt. Im Vergleich zur lokalen Vertrauensbildung sind nicht nur Verhaltensinformationen sondern auch Beweismittel das Ergebnis einer Transaktion. Empfehlungen werden aufgrund dieser Beweismittel für andere Einheiten ausgestellt und von anderen Einheiten entgegengenommen. Außerdem zieht jede Einheit aufgrund der ihr bekannten Beweismittel Schlussfolgerungen. Die Glaubensrevision muss mit diesen zusätzlichen Informationsquellen umgehen können.

Bemerkenswert ist, dass der Einsatz von transaktionalen Beweismitteln keine Auswirkung auf den Glauben, die Vertrauensentscheidungen und die Rahmenbedingungen der Transaktionen hat. Auf diese drei Bereiche werden wir daher in diesem Kapitel nicht weiter eingehen.

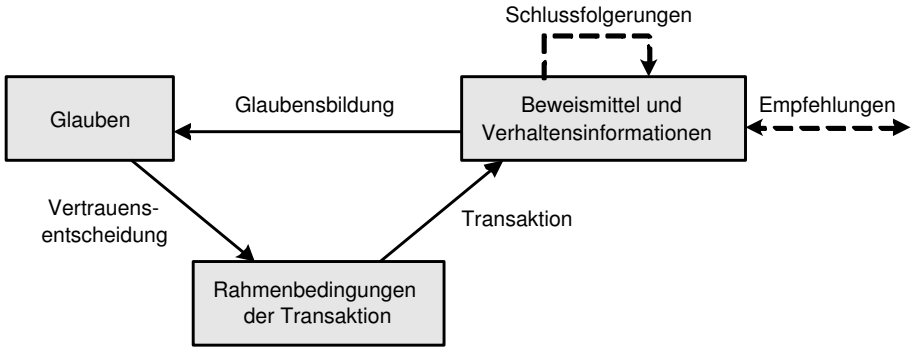

Abbildung 7.3: Datenzentrische Sicht des Kreislaufs der verteilten Vertrauensbildung mit transaktionalen Beweismitteln

7.2 Ausstellen von Beweismitteln in Transaktionen

Die Aussage eines transaktionalen Beweismittels bezieht sich auf das Transaktionsverhalten einer Einheit. Es liegt daher nahe, dass transaktionale Beweismittel im Zuge von Transaktionen ausgestellt werden.

Dieser Abschnitt beschäftigt sich damit, wie diese Idee umzusetzen ist. Zu diesem Zweck identifizieren wir in Abschnitt 7.2.1 zunächst Verträge und Quittungen als zwei Arten von Beweismitteln, deren Ausstellung im Rahmen einer Transaktion sinnvoll ist. Anschließend erweitern wir in Abschnitt 7.2.2 das Zwei-Wege Transaktionsprotokoll um den Austausch dieser Beweismittel. Als Folge davon erhalten wir ein Sechs-Wege Transaktionsprotokoll.

7.2.1 Verträge und Quittungen

Im Folgenden wird definiert, welche Arten von Beweismitteln Verträge und Quittungen sind. Daraus leitet sich ab, wie diese Beweismittel inhaltlich strukturiert sind. Abschließend gehen wir darauf ein, inwiefern das Ausstellen von Verträgen und Quittungen eine Signalisierung darstellt.

Verträge. Im Vorfeld einer Transaktion handeln die beiden Transaktionspartner die Rahmenbedingungen der Transaktion aus, wie sie in Abschnitt 6.4 aufgelistet worden sind. Das Ergebnis der Aushandlung ist dabei nur den Transaktionspartnern bekannt. Es ist daher sinnvoll, dass jeder der beiden Transaktionspartner sich noch vor dem Ausführen der Aktionen darauf festlegt, welche Rahmenbedingungen für die Transaktion ausgehandelt worden sind.

Ein *Vertrag* ist ein Beweismittel, in dem eine solche Festlegung erfolgt. Die Aussage eines Vertrages besteht also darin, dass sich ein Transaktionspartner festlegt, welche Aktion er ausführen wird. Da mehrseitige Festlegungen gemäß dem Modell der Beweismittel nicht möglich sind, müssen beide Transaktionspartner jeweils einen solchen Vertrag ausstellen. Ein Vertrag stellt damit für seinen jeweiligen Aussteller eine einseitige Festlegung dazu dar, welche Verpflichtungen er sich für die kommende Transaktion auferlegt.

Quittungen. Das Ausgang der Transaktion wird dadurch bestimmt, welche Aktionen erfolgreich von den Transaktionspartnern ausgeführt worden sind. Gemäß unserem Systemmodell kennen also nur die Transaktionspartner selbst den Ausgang ihrer Transaktion. Analog zu der Überlegung bei Verträgen ist es damit wünschenswert, dass jeder der beiden Transaktionspartner sich

nach dem Ausführen der Aktionen darauf festlegt, welchen Ausgang die Transaktion gefunden hat.

Eine *Quittung* ist ein Beweismittel, in dem eine solche Festlegung geschieht. Die Aussage einer Quittung beschränkt sich darauf, ob der eigene Transaktionspartner seine Aktion wie versprochen ausgeführt hat. Damit bestätigen sich die Transaktionspartner in ihren Quittungen gegenseitig, dass sie ihre Aktionen ausgeführt haben.

Verträge und Quittungen stehen in einem engen Zusammenhang: Zunächst legt sich Einheit A vor der Transaktion in ihrem Vertrag auf die Aktion, die sie für Einheit B ausführen wird, fest. Nach der Transaktion bestätigt Einheit B durch das Ausstellen einer Quittung, dass die Aktion wie im Vertrag versprochen tatsächlich ausgeführt worden ist. Durch den Erhalt dieser Quittung ist Einheit A also von ihrem Versprechen, das sie in ihrem Vertrag gegeben hat, entbunden. Umgekehrt gilt diese Überlegung auch für den Vertrag der Einheit B und die Quittung der Einheit A.

Das Ausstellen einer Quittung stellt für eine Einheit den letzten Schritt der Transaktion dar. Auf den ersten Blick erscheint es wünschenswert, dass sich die Transaktionspartner das Ausstellen der Quittungen jeweils gegenseitig bestätigen. Dann ergibt sich allerdings die Frage, wie das Ausstellen dieser Bestätigungen wiederum zu quittieren ist. In Analogie zum Problem des koordinierten Angriffs zeigt sich also, dass auf der Grundlage einseitiger Festlegungen kein gemeinsames Wissen über den Ausgang der Transaktion entstehen kann. Daher ist es sinnvoll, das Ausstellen von Quittungen selbst nicht in weiteren Beweismitteln zu bestätigen.

Eine anderer Punkt, der Erklärung bedarf, ist, dass Quittungen immer das erfolgreiche Ausführen der versprochenen Aktion bestätigen. Es sind also insbesondere keine negativen Quittungen vorgesehen, die dem Transaktionspartner Fehlverhalten bescheinigen. Solche negativen Quittungen wären nämlich negative Beweismittel, die gemäß der Einschränkungen aus Abschnitt 7.1.1 nur an unbeteiligte Einheiten weiter gegeben werden können. In diesem Abschnitt beschäftigen wir uns nur mit solchen Beweismitteln, die dem Transaktionspartner direkt im Zuge einer Transaktion ausgestellt und übergeben werden. Im Rahmen des Empfehlungssystems aus Abschnitt 7.3 werden wir allerdings auf den Mangel negativer Quittungen zurückkommen und zu diesem Zweck negative Empfehlungen einführen.

Inhalt von Verträgen und Quittungen. Im Folgenden untersuchen wir, wie die Aussagen von Verträgen und Quittungen inhaltlich zu strukturieren ist. Offensichtlich ist, dass diese Beweismittel eine Kennzeichnung darüber beinhalten müssen, ob es sich um einen Vertrag oder eine Quittung handelt. Darüber hinaus besitzt ein Vertrag folgende Einträge:

- *Identifikator der Transaktion:* Jede Einheit führt eine fortlaufende Nummerierung der Transaktionen, an denen sie teilnimmt. Stellt eine Einheit eine Quittung aus, so fügt sie diese Nummer als Identifikator der Transaktion hinzu. Er dient dazu, einen eindeutigen Zusammenhang zwischen Verträgen und dazugehörigen Quittungen herzustellen.

- *Transaktionspartner:* Hierbei handelt es sich um die Identität des eigenen Transaktionspartners. Sie ist anzugeben, damit im Rahmen des Empfehlungssystems erkennbar ist, wem gegenüber sich eine Einheit in dem Vertrag festgelegt hat.

- *Beschreibung der versprochenen Aktion:* Der Aussteller eines Vertrages verspricht das Ausführen einer Aktion für den Transaktionspartner. Die Beschreibung dieser Aktion gibt einen Hinweis darauf, in welchem Kontext sich die Transaktion bewegt. Im Vertrag muss die-

se Beschreibung enthalten sein, um den Transaktionskontext anderen Einheiten gegenüber deutlich zu machen.

- *Zeitraum der Gültigkeit:* Dieser Eintrag hält fest, bis zu welchem Zeitpunkt der Vertrag gegen seinen Aussteller im Rahmen des Empfehlungssystems verwendet werden kann. Eine Beschränkung der Gültigkeitsdauer von Verträgen ist notwendig, um die erforderliche Speicherausstattung der am Informationssystem teilnehmenden Geräte zu beschränken. Solange ein Vertrag nämlich gültig ist, darf sein Aussteller die zugehörige Quittung nicht löschen, da er sonst im Rahmen des Empfehlungssystems von seinem Transaktionspartner angreifbar ist. Bei der Angabe des Zeitraums der Gültigkeit sind zwei Punkte zu beachten:
  - *Granularität:* In vielen Anwendungsbereichen wie dem Campus-Szenario reicht es aus, den Zeitraum der Gültigkeit durch ein Verfalldatum anzugeben. Dies liegt daran, dass die Zahl der an einem Tage erhaltenen Quittungen im Vergleich zu der Speicherkapazität der Geräte sehr gering ist[1]. Als Folge davon müssen die Geräte nicht über genau synchronisierte Uhren verfügen. Es reicht aus, wenn sie das aktuelle Datum kennen.
  - *Aushandlung:* Der Zeitraum der Gültigkeit eines Vertrags wird im Zuge der Aushandlung der Rahmenbedingungen der Transaktion bestimmt. Die Einheiten müssen dabei abwägen zwischen dem Speicheraufwand bei einem langem Zeitraum und der beschränkten Anwendbarkeitsdauer bei einem kurzen Zeitraum.

Quittungen besitzen dieselbe inhaltliche Struktur wie Verträge. Allerdings ändert sich bei ihnen der Bezug der Einträge. Dies erklärt sich daraus, dass die Quittung einer Einheit sich auf den Vertrag ihres Transaktionspartners bezieht. Daher stimmt der in der Quittung angegebene Identifikator der Transaktion, die Beschreibung der Aktion und der Zeitraum der Gültigkeit mit den Einträgen des zugehörigen Vertrages überein.

Signalisierung mit Verträgen und Quittungen. Dadurch, dass eine Einheit einen Vertrag oder eine Quittung ausstellt, legt sie sich ihrem Transaktionspartner gegenüber auf eine Aussage fest. Damit stellt das Ausstellen dieser transaktionalen Beweismittel eine direkte Signalisierung dar. Dabei unterscheiden sich Verträge und Quittungen darin, was durch sie signalisiert wird:

- *Vertrag:* Der Aussteller signalisiert, dass er beabsichtigt, die Aktion wie versprochen auszuführen. Die Kann-Nicht-Bedingung der glaubwürdigen Signalisierung stellt sich bei Verträgen also wie folgt dar: Kommt es dazu, dass der Aussteller des Vertrages seine Aktion nicht wie versprochen ausführt, dann ist er im Rahmen des Empfehlungssystems von seinem Transaktionspartner unter Zuhilfenahme des Vertrages angreifbar. Stellt eine Einheit also einen Vertrag ohne die Absicht kooperativen Verhaltens in der Transaktion aus, so setzt sie sich dem Risiko aus, durch entsprechende Empfehlungen des Transaktionspartners diskreditiert zu werden.

- *Quittung:* Durch das Ausstellen einer Quittung signalisiert eine Einheit, dass sie ihren Transaktionspartner nicht durch Empfehlungen, die auf seinem Vertrag basieren, diskreditieren wird. Wenn eine Einheit trotz Ausstellen einer Quittung ihren Transaktionspartner diskreditiert, so ist dieser in der Lage, durch Vorzeigen der Quittung diese Diskreditierung zu widerlegen. Durch eine solche Signalstörung diskreditiert sich eine Einheit also letztendlich selbst. In dieser Hinsicht wird die Kann-Nicht-Bedingung erfüllt.

[1] Eine Quittung besitzt eine Größe weit unter einem KiloByte, während der Speicher heute verfügbarer Informationsgeräte selbst bei PDAs mehrere MegaByte umfasst.

Es zeigt sich also, dass die Glaubwürdigkeit der direkten Signalisierung, die durch das Ausstellen von Verträgen und Quittungen erfolgt, eng an das Empfehlungssystem gekoppelt ist. Auf das Verhalten strategischer Einheiten im Hinblick auf das Ausstellen von Verträgen und Quittungen werden wir daher erst in Abschnitt 7.6.1 nach der Besprechung des Empfehlungssystems eingehen.

7.2.2 Sechs-Wege Transaktionsprotokoll

In Abschnitt 6.2 haben wir das Zwei-Wege Transaktionsprotokoll kennen gelernt. Es schreibt vor, in welcher Reihenfolge die Aktionen der beiden Transaktionspartner ausgeführt werden. Dieses Protokoll wird durch das Einbeziehen des Austausches von Verträgen und Quittungen zum *Sechs-Wege Transaktionsprotokoll* erweitert, das in Abbildung 7.4 dargestellt ist. Die Schritte des Protokolls stellen sich wie folgt dar:

1. Einheit A stellt einen Vertrag darüber aus, dass sie ihre Aktion für Einheit B ausführen wird. Diesen Vertrag übermittelt sie der Einheit B. Diese überprüft, ob der Inhalt des Vertrags dem entspricht, was die beiden Transaktionspartner zuvor für die Rahmenbedingungen der Transaktion ausgehandelt haben.

2. Entspricht der Vertrag den Erwartungen von Einheit B, so stellt sie ihrerseits der Einheit A gegenüber einen Vertrag aus. Dieser wird von Einheit A ebenso auf seine Übereinstimmung mit dem Ergebnis der Aushandlung überprüft.

3. Ist auch der Vertrag der Einheit B korrekt, so führt Einheit A ihre Aktion aus. Wie im Zwei-Wege Protokoll wird die Aktionsausführung von Einheit B auf ihre Richtigkeit überprüft.

4. Einheit B führt ihrerseits ihre Aktion für Einheit A aus. Auch die Richtigkeit dieser Aktionsausführung wird überprüft.

5. Einheit A stellt eine Quittung über die erfolgreiche Ausführung der Aktion durch Einheit B aus und übergibt sie an dieselbe. Einheit B überprüft daraufhin, ob die Quittung sich tatsächlich auf den Vertrag bezieht, den sie in Schritt 2 ausgestellt hat.

6. Entspricht die Quittung den Erwartungen der Einheit B, so erwidert sie dies ihrerseits mit einer entsprechenden Quittung über die Aktionsausführung von Einheit A. Auch diese Quittung wird von Einheit A auf ihre Richtigkeit überprüft.

Das Sechs-Wege Transaktionsprotokoll stimmt weitgehend mit dem optimistischen Austauschprotokoll für schwache Fairness aus Abschnitt 2.3.2 überein. Der einzige substantielle Unterschied gegenüber diesem Protokoll liegt darin, dass die Quittung der Einheit B in einem separaten sechsten Schritt ausgestellt wird. Damit wird sichergestellt, dass beide Transaktionspartner beim Austausch der Verträge, Aktionen und Quittungen jeweils dieselbe Position bezüglich des Betrugsrisikos einnehmen.

Als Ergebnis der Transaktion stehen beiden Einheiten außer den erhaltenen Beweismitteln die Beobachtungen darüber zur Verfügung, wie sich ihr jeweiliger Transaktionspartner verhalten hat. Gemäß Abschnitt 6.2 ist eine solche Verhaltensinformation entweder die Kooperation oder der Betrug des Transaktionspartners. Um als kooperativ aufgefasst zu werden, muss eine Einheit nicht nur ihre Aktion ausführen. Darüber hinaus hat sie einen entsprechenden Vertrag und eine Quittung auszustellen. Unter Betrug verstehen wir damit auch das Nichtausstellen eines Vertrags

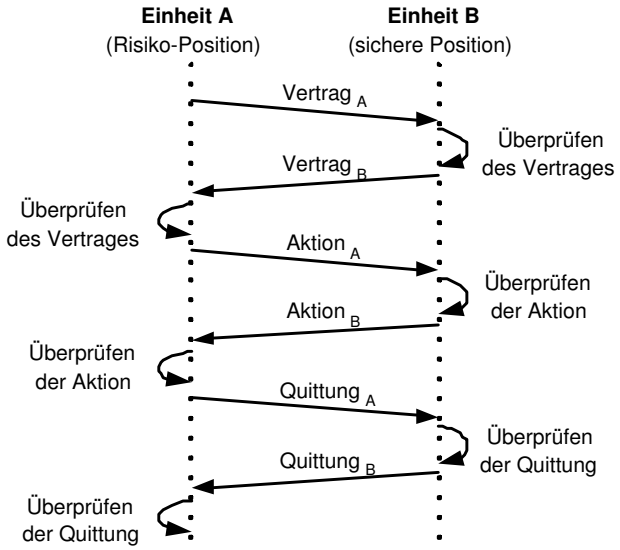

Abbildung 7.4: Das Sechs-Wege Transaktionsprotokoll

oder einer Quittung. Die erste Norm des Systementwurfs (siehe Seite 153) wird damit auf alle Schritte des Sechs-Wege Transaktionsprotokolls ausgedehnt.

Der logische Zusammenhang zwischen den Schritten des Sechs-Wege Transaktionsprotokolls und der Aushandlungsphase, die der Transaktion vorgelagert ist, wird in Abbildung 7.5 verdeutlicht. Das Ergebnis der Aushandlung ist zwar beiden Transaktionspartnern bekannt. Es wird aber nur durch den Austausch von Verträgen, die dieses Ergebnis festhalten, auch anderen Einheiten gegenüber glaubhaft vermittelbar. In diesem Sinne besitzt die Festlegung auf die Rahmenbedingung der Transaktion in Verträgen das Potential, dass alle interessierte Einheiten von dem Ergebnis der Aushandlung erfahren. In der Abbildung wird dies mit globalem Wissen bezeichnet. Hingegen ist das Ergebnis der Aushandlung ohne Vertragsaustausch nur den beiden Transaktionspartner bekannt. Ob die Versprechen, auf das sich die Transaktionspartner gegenseitig festgelegt haben, eingehalten werden, entscheidet sich in der Phase der Aktionsausführung. Am Ende dieser Phase wissen zwar die Transaktionspartner, wer sein Versprechen eingehalten hat. Um dieses Wissen global zugänglich zu machen, bedarf es jedoch zusätzlich des Austausches der Quittungen. Die Form der Darstellung, die in der Abbildung gewählt ist, verdeutlicht den inhaltlichen Zusammenhang der Phasen der Transaktion: Die Kooperation zwischen zwei Transaktionspartner besteht im Aushandeln von gegenseitigen Versprechen, über deren Einhaltung der Austausch der Aktionen entscheidet. Da sowohl die Versprechen als auch ihre Einhaltung nur den beiden Transaktionspartnern bekannt ist, bedarf es des Einsatzes von Verträgen und Quittungen. Diese Beweismittel ermöglichen, dass jede andere Einheit von den Versprechen und ihrer Einhaltung erfahren kann.

7.3 Beweismittel in Empfehlungen

Der Einsatz von Beweismittel hat einen maßgeblichen Einfluss auf die Gestaltung des Empfehlungssystems: Damit eine Einheit empfehlen kann, müssen ihr entsprechende Festlegungen anderer Einheiten in Form von Beweismitteln vorliegen. Daher bestimmt die Art der eingesetzten Beweismittel, welche prinzipiellen Empfehlungsmöglichkeiten gegeben sind.

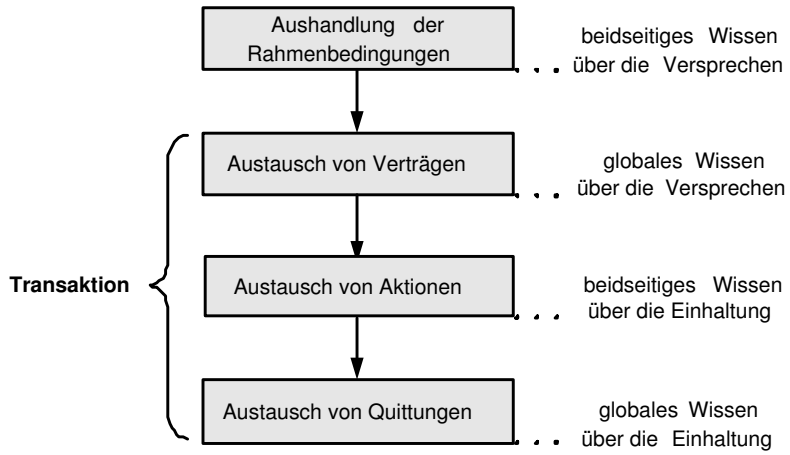

Abbildung 7.5: Logischer Zusammenhang zwischen den Phasen einer Transaktion

In diesem Abschnitt widmen wir uns also der Frage, wie das Empfehlungssystem auf der Grundlage transaktionaler Beweismittel zu entwerfen ist. Hierfür gehen wir in Abschnitt 7.3.1 darauf ein, welche Arten von Empfehlungen sich ergeben. Die Vorschriften bezüglich des Einholens und Ausstellens solcher Empfehlungen werden in Abschnitt 7.3.2 vorgestellt.

7.3.1 Empfehlungsarten

Vor der Beschreibung der einzelnen Arten von Empfehlungen gehen wir im Folgenden zunächst auf die Prinzipien ein, die dem Entwurf des Empfehlungssystems zugrunde liegen. Sie leiten die Definition der drei Arten von Empfehlungen, der wir uns anschließend zuwenden. Bei ihnen handelt es sich um negative Empfehlungen, Selbstempfehlungen und Typbeweise. Diese Empfehlungsarten werden abschließend in einem Beispiel untereinander in Beziehung gebracht.

Prinzipien. Der Einsatz von Empfehlungen ist nur dann sinnvoll, wenn durch sie gewisse Sachverhalte glaubhaft signalisiert werden können. Hierbei gibt es zwei prinzipielle Möglichkeiten: **(1)** Zur Untermauerung der in der Empfehlung getroffenen Aussage wird ein Beweismittel angeführt, das eine zugehörige Festlegung einer anderen Einheit enthält. In diesem Fall handelt es sich beim Empfehlen um indirektes Signalisieren. **(2)** Die Aussage der Empfehlung kann nur dann hinreichend untermauert werden, wenn sich der Empfehler in einem entsprechenden Beweismittel auf sie festlegt. Dies entspricht dem direkten Signalisieren.

Beide Möglichkeiten schließen sich nicht gegenseitig aus. Eine Empfehlung kann gleichzeitig Beweismittel Anderer enthalten und selbst ein Beweismittel sein. Dies trifft auf die nachfolgend eingeführte negative Empfehlung zu.

Das Empfehlungssystem übernimmt eine weitere Aufgabe. Im Zuge der direkten und indirekten Signalisierung muss der Empfehler entsprechende Beweismittel bereitstellen, um den jeweiligen Empfänger der Empfehlung überzeugen zu können. Dadurch kommt es zu einer Verteilung und Weitergabe der Beweismittel zwischen den Einheiten.

Bei jeder der nachfolgend vorgestellten Empfehlungsarten sind folglich zwei Fragen zu klären:

1. Handelt es sich um eine direkte oder indirekte Signalisierung (oder um Beides)?

2. Für den Fall der indirekten Signalisierung: Welche Beweismittel Anderer werden zur Unterstützung der Empfehlung weiter gegeben?

Negative Empfehlungen. Bei der Einführung von Quittungen ist in Abschnitt 7.2.1 gezeigt worden, dass Quittungen stets positive Aussagen beinhalten. Wenn eine Einheit betrügendes Verhalten ihres Transaktionspartners wahrnimmt, wird sie ihm gegenüber also dies nicht in einer entsprechenden negativen Quittung bescheinigen. Festlegungen zu negative Aussagen (wie der des Betrugs) können nur an unbeteiligte Einheiten weiter gegeben werden.

Eben hierin liegt der Zweck einer negativen Empfehlung: Macht eine Einheit eine negative Transaktionserfahrung, so stellt sie eine negative Empfehlung über ihren Transaktionspartner aus. Diese negative Empfehlung wird nicht an den Transaktionspartner selbst sondern an unbeteiligte Einheiten weiter gegeben, die ihr Interesse an einer solchen negativen Empfehlungen bekundet haben. So wird sichergestellt, dass die Einheiten sich auch über ihre negativen Transaktionserfahrungen berichten können.

Welche Art der Signalisierung stellt eine negative Empfehlung dar? Zunächst ist der Empfänger der Empfehlung davon zu überzeugen, dass die Transaktion, auf die sich die Empfehlung bezieht, tatsächlich stattgefunden hat. Zu diesem Zweck muss der Vertrag des angeblich betrügenden Transaktionspartners vorgezeigt werden. In dieser Hinsicht handelt es sich um eine indirekte Signalisierung. Andererseits fordern wir, dass sich der Aussteller einer negativen Empfehlung wie bei einer Quittung zu seiner Aussage mit Hilfe eines entsprechenden Beweismittels festlegt. Dies bedeutet, dass eine negative Empfehlung selbst ein Beweismittel ist. Damit erhalten wir neben Verträgen und Quittungen eine dritte Art transaktionaler Beweismittel. Das Ausstellen einer negativen Empfehlung stellt also eine direkte Signalisierung dar.

Die Forderung, dass der Aussteller einer negativen Empfehlung sich zu seiner Aussage festlegt, besitzt zwei weit reichende Vorteile:

- *Möglichkeit der Weitergabe:* Negative Empfehlungen können glaubhaft weiter gegeben werden. Hierfür reicht es aus, dass sich eine Einheit beim Empfang einer negativen Empfehlung die Empfehlung bei sich lokal ablegt und auf Anfrage auch anderen Einheiten zur Verfügung stellt. Diese Weitergabe erfolgt ohne Verminderung der Glaubwürdigkeit der negativen Empfehlung, da sie als Beweismittel ausgestellt und weitergeleitet wird. Besonders in Informationssystemen wie im Campus-Szenario ist diese Möglichkeit der Weitergabe negativer Empfehlungen von Vorteil: Da die Einheiten über ein Ad-hoc Netz miteinander kommunizieren, ist für jede Einheit nur ein kleiner Bruchteil der anderen Einheiten erreichbar. Mit ihrer negativen Empfehlung kann eine Einheit daher nur dann alle interessierten Einheiten erreichen, wenn diese Empfehlung von anderen Einheiten weiter gegeben werden kann.

- *Vergleich der Festlegungen:* Eine Einheit könnte den Versuch unternehmen, nach dem Ausstellen einer Quittung ihren Transaktionspartner anderen Einheiten gegenüber zu diffamieren. Hierfür muss sie eine negative Empfehlung bezüglich dieser Transaktion ausstellen. Dadurch stehen die Aussagen der Quittung und der negativen Empfehlung in einem direkten Widerspruch. Es ist wünschenswert, dass andere Einheiten diesen Widerspruch erkennen, damit sie ihren Glauben über diese Einheit, die sich durch diese Diffamierung *inkonsistent* verhalten hat, entsprechend revidieren können. Ein solches Erkennen wird dadurch ermöglicht, dass die Quittung und die negative Empfehlung als Beweismittel ausgestellt werden. Durch den Vergleich dieser Beweismittel ist jede Einheit in der Lage, die widersprüchlichen Festlegungen der diffamierenden Einheit zu erkennen. In dieser Hinsicht löst

der Einsatz von Beweismitteln das *Problem byzantinischer Generäle* aus Abschnitt 4.2: Nur durch die Verwendung nicht-abstreitbarer Festlegungen lässt sich diejenige Einheit identifizieren, die sich inkonsistent verhält.

Der zweite Punkt macht darauf aufmerksam, dass durch die Verwendung von Beweismitteln inkonsistentes Empfehlungsverhalten erkennbar wird. Diese Beobachtung nützen wir aus, indem wir diesbezüglich die zweite Norm des Systementwurfs formulieren:

2. Norm:
Eine Einheit darf sich unter keinen Umständen zu Aussagen festlegen, die sich gegenseitig widersprechen.

Bei dieser Norm sind die beiden Anforderungen an Normen aus Abschnitt 5.4.3 erfüllt: Die Norm schreibt in dem Sinne kooperatives Verhalten vor, als sich die Einheiten in ihren Festlegungen nicht gegenseitig in die Irre führen dürfen. Außerdem kann das Norm-bezogene Verhalten kontrolliert werden, da Festlegungen durch den Einsatz von Beweismitteln nicht-abstreitbar sind.

Ein entscheidender Vorteil dieser Norm über die Norm bezüglich des Transaktionsverhaltens ist, dass unbeabsichtigtes Fehlverhalten ausgeschlossen werden kann. Dies spiegelt sich in der Definition der Norm durch die Wortgruppe *unter keinen Umständen* wider. Eine normative Einheit wird also nur dann eine negative Empfehlung ausstellen, wenn sie sich sicher ist, dass sie zuvor keine Quittung in der jeweiligen Transaktion übergeben hat.

Bei den eingeführten Arten von transaktionalen Beweismitteln gibt es nur eine Möglichkeit, dass es zu einem Widerspruch zwischen Festlegungen kommt, nämlich das Ausstellen einer Quittung und einer negativen Empfehlung über dieselbe Transaktion. Die beiden nachfolgend vorgestellten Empfehlungsarten beschäftigen sich damit, die Erkennbarkeit eines solchen inkonsistenten Verhaltens weiter zu erhöhen.

Selbstempfehlungen. Durch die Einführung von negativen Empfehlungen wird es möglich, dass eine Einheit von ihrem ehemaligen Transaktionspartner anderen Einheiten gegenüber diskreditiert wird. Diese Möglichkeit besteht auch dann, wenn der Transaktionspartner zuvor der Einheit kooperatives Verhalten durch die Ausstellung einer entsprechenden Quittung bescheinigt hat. Um sich gegen solche Diskreditierungen zu schützen, bietet es sich daher für eine Einheit an, die negative Empfehlung durch das Vorzeigen dieser Quittung zu widerlegen.

Eben hierin liegt der Zweck einer Selbstempfehlung: Jede Einheit ist in der Lage, die Quittungen, die sie von ihren ehemaligen Transaktionspartnern erhalten hat, in einer Empfehlung zusammenzustellen. Bei einer solchen Empfehlung reden wir im Folgenden von einer Selbstempfehlung. Dadurch, dass Selbstempfehlungen von den begleitenden Quittungen gestützt werden, handelt es sich bei ihnen um eine Möglichkeit der indirekten Signalisierung, die zur Verbreitung von Quittungen führt. In dieser Hinsicht ergänzen Selbstempfehlungen die negativen Empfehlungen, deren Ausstellung zur Verteilung der Verträge führt.

Durch die Einsatz von Selbstempfehlungen werden somit zwei Ziele erreicht:

- *Sicht des Empfehlers:* Der Aussteller einer Selbstempfehlung stellt sicher, dass er vor etwaigen negativen Empfehlungen seiner ehemaligen Transaktionspartner geschützt ist.

- *Sicht des Gesamtsystems:* Durch die Verbreitung der Quittungen ergibt sich überhaupt erst die Möglichkeit, dass das inkonsistente Empfehlungsverhalten einer Einheit aufgedeckt wird. Dies lässt sich an folgendem Beispiel festmachen: Nehmen wir an, dass Einheit A und B

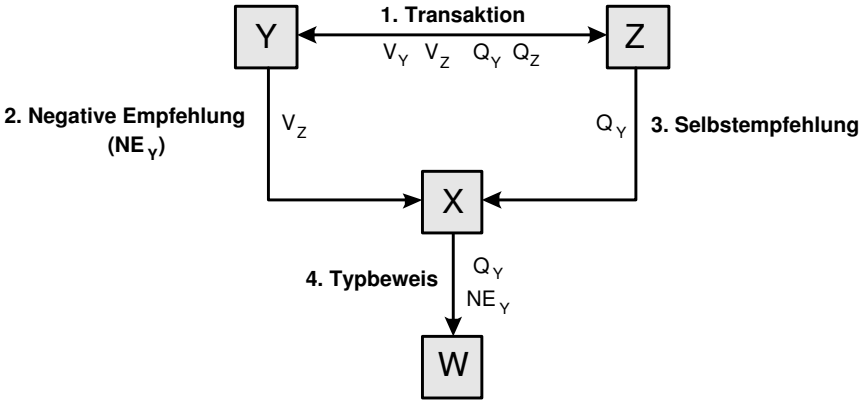

Abbildung 7.6: Beispiel für das Zusammenwirken der Empfehlungsarten

eine erfolgreiche Transaktion durchgeführt haben, in der es zum Austausch entsprechender Quittungen gekommen ist. Dennoch stellt Einheit A eine negative Empfehlung bezüglich dieser Transaktion an Einheit C aus. Zu diesem Zeitpunkt kann Einheit C nicht erkennen, dass Einheit A mit der zweiten Norm gebrochen hat. Dies ändert sich allerdings, wenn Einheit B im Zuge einer Selbstempfehlung an Einheit C die entsprechende Quittung übergibt. In diesem Moment ist Einheit C in der Lage, die beiden erhaltenen Festlegungen, die Quittung und die negative Empfehlung, in Zusammenhang zu bringen und ihren Widerspruch zu erkennen. Nur dadurch kann Einheit C ihren Glauben über die sich inkonsistent verhaltende Einheit A revidieren. Aus der Sicht des Gesamtsystems haben Selbstempfehlungen also die Aufgabe, die Erkennung inkonsistenten Verhaltens und damit die Kontrolle über Norm-bezogenes Verhalten zu ermöglichen.

Typbeweise. Erkennt eine Einheit, dass sich eine andere Einheit zu inkonsistenten Aussagen festgelegt hat, so ist sie nicht nur in der Lage, ihren Glauben über diese Einheit zu revidieren. Darüber hinaus kann sie jede beliebige Einheit von der Inkonsistenz dieser Einheit überzeugen, indem sie die im Widerspruch stehenden Beweismittel vorlegt. Dabei kann nicht nur auf das inkonsistente Verhalten der Einheit sondern auch auf ihren Typ geschlossen werden. Da sich normative Einheiten niemals zu inkonsistenten Aussagen festlegen, muss es sich bei dieser Einheit um eine strategische handeln.

Ein Typbeweis ist eine Empfehlung, die aufzeigt und belegt, von welchem Typ der Empfohlene ist. Die einzig mögliche Form eines Typbeweises besteht also darin, dass widersprüchliche Festlegungen des Empfohlenen vorgezeigt werden und daraus gefolgert wird, dass der Empfohlene strategisch ist. Auf der Grundlage der transaktionalen Beweismittel bedeutet dies, dass ein Typbeweis aus einem Paar von Quittung und negativer Empfehlung besteht, die sich gegenseitig widersprechen. Das Ausstellen eines Typbeweises stellt damit eine indirekte Signalisierung dar, die zur Verteilung von Quittungen, negativen Empfehlungen und letztlich dem Wissen um die Inkonsistenz einer Einheit führt. Daher wird durch den Einsatz von Typbeweisen die Kontrolle über die Einhaltung der zweiten Norm weiter verstärkt.

Beispiel für das Zusammenwirken der Empfehlungsarten. Um die in diesem Abschnitt vorgestellten Empfehlungsarten untereinander in Beziehung zu setzen wird im Folgenden ein beispielhafter Ablauf beschrieben. Eine Übersicht der Schritte des Ablaufs gibt Abbildung 7.6:

1. Einheit Y und Einheit Z nehmen gemeinsam an einer Transaktion teil, die erfolgreich abgeschlossen wird. Am Ende der Transaktion hat somit Einheit Y den Vertrag V_Z und die Quittung Q_Z der Einheit Z erhalten und umgekehrt.

2. Einheit Y stellt eine negative Empfehlung NE_Y über Einheit Z aus und übermittelt sie an die Einheit X. Damit die negative Empfehlung gültig ist, muss der Vertrag V_Z an sie angehängt sein. Damit wird auch bestimmt, dass sich die negative Empfehlung auf die Transaktion aus Schritt **(1)** bezieht.

3. Einheit Z stellt eine Selbstempfehlung aus und übergibt sie an Einheit X. Sie besteht aus der Quittung Q_Y, die in der vorigen Transaktion von Einheit Y erhalten worden ist.

4. Einheit X erkennt, dass Einheit Y inkonsistent gehandelt hat und daher strategisch ist. Damit ist sie in der Lage, einen entsprechenden Typbeweis an eine andere Einheit W zu schicken. Dieser Typbeweis enthält die beiden sich widersprechenden Beweismittel der Einheit Y, nämlich ihre Quittung Q_Y und negative Empfehlung NE_Y. Als Folge davon ist auch für Einheit W nachvollziehbar, dass Einheit Y strategisch ist.

Bei diesem Ablauf könnten die Schritte 2 und 3 auch in einer umgekehrten Reihenfolge stehen. Entscheidend ist, dass die Inkonsistenz einer Einheit immer dann erkannt wird, wenn je eine ihrer Quittungen und negative Empfehlungen, die sich auf dieselbe Transaktion beziehen, aufeinander treffen. Weiterhin bleibt zu bemerken, dass nach dem vierten Schritt die Einheiten X und W jeder weiteren Einheit, insbesondere der Einheit Z, einen Beweis über den Typ der Einheit Y liefern können. Daraus wird ersichtlich, dass inkonsistentes Handeln systemweit erkannt und durch entsprechende lokale Glaubensrevisionen bestraft werden kann.

7.3.2 Vorschriften über das Empfehlungsverhalten

In diesem Abschnitt beschäftigen wir uns mit den Vorschriften, die den Umgang mit den drei Arten von Empfehlungen regeln. Dabei unterscheiden wir zwischen den zwei Rollen, die eine Einheit bezüglich einer Empfehlung spielen kann. Um Informationen über eine andere Einheit zu erhalten, müssen Empfehlungen *eingeholt* werden. Dies geschieht durch eine entsprechende Anfrage an eine andere Einheit. Bei Erhalt einer solchen Anfrage muss sich eine Einheit entscheiden, wie sie eine entsprechende Empfehlung *zusammenstellt*. Auf diese beiden Rollen gehen wir im Folgenden nacheinander ein.

Einholen von Empfehlungen. Vorschriften für das Einholen von Empfehlungen müssen die vier folgenden Fragen beantworten:

- *Über wen:* Über welche Einheit sollen Informationen eingeholt werden?
- *Wann:* Unter welchen Umständen wird eine Empfehlung eingeholt?
- *Was:* Welche Art der Empfehlung wird eingeholt?
- *Von wem:* Welche Einheit wird angefragt, eine entsprechende Empfehlung auszustellen?

Für die ersten beiden Punkte können wir eine klare Richtlinie angeben: Empfehlungen werden immer über solche Einheiten angefordert, die ein potentieller Partner in einer anstehenden Transaktion sind. Das Einholen von Empfehlungen zielt also darauf, eine erweiterte Grundlage für das

Treffen von Vertrauensentscheidungen bezüglich möglicher Transaktionspartner zu bilden. Eben hierin liegt der eigentliche Zweck des Empfehlungssystems. Ein weiterer Vorteil dieser Verfahrensweise liegt darin, dass eine Einheit vor dem Hintergrund einer bevorstehenden Transaktion einen besonderen Anreiz erhält, sich selbst zu empfehlen. Sie erhofft sich dadurch, die Vertrauensentscheidung ihres potentiellen Transaktionspartners zu ihren Gunsten zu beeinflussen.

Die Frage nach der Art der Empfehlung lässt sich mit Hilfe eines Konzepts aus der Entscheidungstheorie beantworten [Ber85]. Zusätzliche Informationen für das Treffen von Entscheidungen sind umso wertvoller, je wahrscheinlicher sie die Entscheidung beeinflussen. Bei einer Vertrauensentscheidung stehen nur zwei Alternativen zur Verfügung. Entweder ist eine Einheit bereit zu der gemeinsamen Transaktion oder sie ist es nicht. Daher müssen wir uns beim Einholen von Empfehlungen fragen, inwiefern sich eine Tendenz für oder wider das Eingehen einer Transaktion umkehren lässt. Ein Indikator ist hierfür die Höhe des erwarteten Nutzens, der sich aus der Teilnahme an der Transaktion ergibt und in Abschnitt 6.4 angegeben wurde. Liegt dieser Wert nahe null, so ist es am wahrscheinlichsten, dass zusätzliche Information zu einer Änderung der Vertrauensentscheidung führt. In diesem Fall ist das Einholen von Empfehlungen besonders vorteilhaft. Diese Aussage lässt sich bezüglich der Art der eingeholten Empfehlung folgendermaßen differenzieren: Selbstempfehlungen sind eher dann von Interesse, wenn der erwartete Nutzen der Transaktion nahe null oder leicht negativ ist. Dies liegt daran, dass gemäß der Vorschriften der Glaubensrevision aus dem nachfolgenden Abschnitt 7.4 Selbstempfehlungen zu einer Aufwertung einer Einheit führen können. Umgekehrt erhalten wir aus einer analogen Überlegung, dass negative Empfehlungen und Typbeweise dann einzuholen sind, wenn der erwartete Nutzen nahe null oder leicht positiv ist.

Die letzte zu klärende Frage betrifft die Wahl der Einheit, von der eine Empfehlung eingeholt werden soll. Bei einer Selbstempfehlung beantwortet sich diese Frage von selbst. Bei den anderen beiden Empfehlungsarten stellt sich der Sachverhalt schwieriger dar: Negative Empfehlungen und Typbeweise über einen potentiellen Transaktionspartner lassen sich nur vor den Einheiten anfordern, die sich in Kommunikationsreichweite befinden. Stehen mehrere solche Einheiten zur Verfügung so ist einzuschätzen, welche Einheit am Ehesten eine entscheidende Empfehlung ausstellen kann. Bei Typbeweisen gibt es hierfür keinen Anhaltspunkt, so dass sie zufällig von einer erreichbaren Einheit anzufordern sind. Die Situation sieht bei negativen Empfehlungen jedoch anders aus. Gemäß den Vorschriften aus dem nachfolgenden Abschnitt 7.4 findet eine negative Empfehlung bei der Revision des Typglaubens ein umso größeres Gewicht, je normativer der Empfehler wahrgenommen wird. Daher sind negative Empfehlungen bevorzugt von den normativer erscheinenden Einheiten einzuholen. Zwei mögliche Algorithmen für die Bestimmung des Einholens negativer Empfehlungen gestalten sich daher wie folgt:

- *Rein glaubensbasierter Algorithmus:* Die erreichbaren Einheiten werden in einer Liste nach fallendem Glauben über ihre Normativität angeordnet. Bei der Entscheidung, welche Einheit angefragt wird, wird diese Liste von oben abgearbeitet. Sollen also zum Beispiel drei Einheiten nach negativen Empfehlungen angefragt werden, so werden die ersten drei Einheiten dieser Liste gewählt.

- *Gemischter Algorithmus:* Der rein glaubensbasierte Algorithmus kann um ein probabilistisches Element erweitert werden, um auch Zugang zu den Transaktionserfahrungen von weniger normativ erscheinenden Einheiten zu erhalten. Hierzu werden die erreichbaren Einheiten mit einer Wahrscheinlichkeit angefragt, die mit dem Glauben über ihren Typ gewichtet ist.

Zusammenstellen von Empfehlungen. Vor dem Ausstellen einer Empfehlung muss sich eine Einheit entscheiden, mit welchen Beweismitteln sie den Adressaten der Empfehlung am Ehesten überzeugen kann. Der Typbeweis nimmt hierbei eine Sonderrolle ein, da sich bei ihm diese Frage so nicht stellt: Es reicht nachzuweisen, dass sich der Empfohlene ein einziges Mal inkonsistent verhalten hat, um den Empfohlenen als strategische Einheit darzustellen. Daher ist bei der Zusammenstellung eines Typbeweis ein beliebiges Paar von einer sich widersprechenden Quittung und negativer Empfehlung zu wählen, sofern ein solches Paar existiert. Ansonsten kann kein Typbeweis ausgestellt werden.

Bei der Zusammenstellung einer Selbstempfehlung muss eine Einheit Überlegungen darüber anstellen, welcher bisheriger Transaktionspartner sie diffamieren könnte. Zu ihrem eigenen Schutz wird sie die Quittungen solcher ehemaliger Transaktionspartner bevorzugt in der Selbstempfehlung anzuführen. Als Anhaltspunkt dafür, von wem eine Diffamierung droht, könnte zwar der Typglaube über die bisherigen Transaktionspartner genommen werden. Allerdings sind auch die Quittungen von eher normativ erscheinenden Einheiten zu verbreiten, da es sich bei diesen Einheiten durchaus auch um strategische Einheiten handeln kann. Ein einfacher und dennoch sinnvoller Algorithmus für das Zusammenstellen von Selbstempfehlungen besteht also darin, dass unter den Quittungen, die in der Beweismittel-Basis abgelegt sind, zufällig eine bestimmte Zahl ausgewählt und in die Selbstempfehlung einbezogen wird.

Für das Ausstellen von negativen Empfehlungen geben wir eine eindeutige Vorschrift vor: Für alle negativen Transaktionserfahrungen, von denen man dem Adressaten der Empfehlung noch nicht berichtet hat, wird jeweils eine negative Empfehlung ausgestellt. Diese Vorschrift führt dazu, dass normative Einheiten immer negativ empfehlen, soweit sie danach angefragt werden. Dadurch sind sie auf das Ausstellen negativer Empfehlungen vor-festgelegt. Diese auf den ersten Blick rigide erscheinende Vorschrift erfährt ihre Berechtigung aus zwei Gesichtspunkten: **(1)** Durch den Einsatz negativer Empfehlungen verfolgen wir das Ziel, dass sich die Einheiten über ihre Beobachtungen betrügenden Transaktionsverhaltens gegenseitig berichten. Es ist daher wünschenswert, dass alle solchen negativen Transaktionserfahrungen an diejenigen Einheiten weiter gegeben werden, die durch das Stellen entsprechender Anfragen Interesse an ihnen bekundet haben. **(2)** Die Vor-Festlegung normativer Einheit bezüglich ihres Empfehlungsverhaltens führt dazu, dass auch strategische Einheiten wann immer möglich negativ empfehlen, obwohl sie an sich keinen direkten Anreiz dazu haben. Dies wird die Analyse aus Abschnitt 7.6.2 zeigen. Damit ist sichergestellt, dass strategische Einheiten sich an den Kosten der sozialen Kontrolle beteiligen.

Bei der Zusammenstellung von Selbstempfehlungen und negativen Empfehlungen ist ein Kompromiss zwischen zwei gegensätzlichen Ansätzen zu finden:

- *Speicheraufwand:* Jede Einheit merkt sich, welches Beweismittel sie bereits an wen weiter gegeben hat. Bei der Zusammenstellung von Empfehlungen ist eine Einheit somit in der Lage, das wiederholte (und damit unnötige) Versenden von Beweismitteln an dieselbe Einheit zu vermeiden. Sie handelt sich dadurch aber auch einen erheblichen Speicheraufwand von $O(b \cdot e)$ ein, wobei b die Zahl der bekannten Beweismittel und e die Zahl der bekannten Einheiten ist.

- *Kommunikationsaufwand:* Um diesen Speicheraufwand und den damit einher gehenden Verwaltungsaufwand zu vermeiden, könnte eine Einheit auch auf das Gedächtnis der Zusammenstellung ihrer bisherigen Empfehlungen verzichten. Dies bewirkt allerdings, dass bei ihren Empfehlungen auch Beweismittel einbezogen werden, die sie an den Adressaten der Empfehlung bereits zu einem früheren Zeitpunkt geschickt hat. Daraus ergibt sich ein erhöhter Kommunikationsaufwand für die Einheit.

Ein Abgleich zwischen dem Speicher- und Kommunikationsaufwand muss das jeweilige Ressourcenprofil des Gerätes, auf dem sich die Einheit befindet, berücksichtigen. Wie dieser Abgleich im Einzelnen aussieht, ist für unsere weiteren Betrachtungen nicht von Belang. Dies liegt daran, dass die Beweismittel- und Wissensverwaltung aus Abschnitt 7.5 dafür sorgt, dass Beweismittel nicht doppelt berücksichtigt werden können.

7.4 Beweismittel-basierte Glaubensrevision

Beweismittel und Empfehlungen zielen letztendlich darauf, dass die Erfahrungen und Sichtweisen einer Einheit in die Glaubensbildung anderer Einheiten mit einfließen. Dazu ist festzulegen, wie der Glauben einer Einheit bei Erhalt einer Empfehlung zu revidieren ist.

Mit dieser Aufgabe beschäftigen wir uns in diesem Abschnitt. Zu diesem Zweck identifizieren wir zunächst in Abschnitt 7.4.1 die Arten von Ereignissen der Glaubensrevision, zu denen der Erhalt von Empfehlungen führt. Anschließend arbeiten wir für diese Ereignisse die Revisionsvorschriften in Abschnitt 7.4.2 aus.

7.4.1 Arten von Ereignissen

Im Bezug auf die Glaubensbildung verstehen wir unter einem *Ereignis* eine Änderung der Informationslage einer Einheit, die zur Revision ihres Glaubens führen muss. In Abschnitt 6.3.3 haben wir bereits Arten von Ereignissen identifiziert, die sich auf lokal verfügbare Informationen beziehen. Diesen stellen wir im Folgenden weitere Ereignisarten hinzu, die sich auf den Erhalt von Empfehlungen beziehen. Zu diesem Zweck gehen wir für die drei Empfehlungsarten darauf ein, ob und wie sie zu einer Glaubensrevision führen sollten.

Negative Empfehlung. Welchen Informationsgehalt besitzt eine negative Empfehlung, die von einer Einheit über eine andere Einheit ausgestellt wurde? Bei dem Erhalt einer solchen negativen Empfehlung sind zwei Sachverhalte möglich:

- In der Transaktion zwischen diesen beiden Einheiten kam es zu einem *Konflikt*, insofern als einer der beiden Transaktionspartner betrogen hat. Dabei ist für einen Außenstehenden nicht zu entscheiden, welcher der beiden, der Empfehler oder der negativ Empfohlene, betrogen hat.

- Der Empfehler handelt auf inkonsistente Weise, da er dem Empfohlenen in der entsprechenden Transaktion durch das Ausstellen einer Quittung bereits kooperatives Verhalten bestätigt hat.

Dass es keinen dritten Fall gibt, zeigt die folgende Überlegung: Nehmen wir an, dass der Empfehler sich nicht inkonsistent verhält, also keine Quittung ausgestellt hat. Daraus folgern wir, dass das Sechs-Wege Transaktionsprotokoll nicht erfolgreich durchlaufen worden ist. Es gibt also mindestens einen Schritt des Protokolls, der unterlassen worden ist. Damit hat einer der Transaktionspartner betrogen.

Für die Bewertung von negativen Empfehlungen verfolgen wir einen Ansatz, der vom Zutreffen des ersten Sachverhalts ausgeht: Wird eine negative Empfehlung erhalten, so wird daraus gefolgert, dass ein Konflikt zwischen dem Empfehler und Empfohlenen aufgetreten ist. Dies löst ein Revision des eigenen Glaubens aus. Damit ist die erste Ereignisart die Wahrnehmung eines

Konflikts. Ein solches Ereignis tritt immer dann ein, wenn eine negative Empfehlung erhalten wird.

Wie ist mit der Möglichkeit umzugehen, dass in Wirklichkeit kein Konflikt existiert, sich der Aussteller der negativen Empfehlung aber inkonsistent verhalten hat? Zunächst einmal ist festzuhalten, dass sich der inkonsistente Empfehler dadurch allen anderen Einheiten gegenüber als strategisch zu erkennen gibt und daher enorme Folgekosten in Kauf nehmen müsste. Diese Überlegung zeigt zwar, dass inkonsistentes Empfehlungsverhalten nicht im Sinne einer Einheit und daher nicht zu erwarten ist. Allerdings kann solches Verhalten auch nicht kategorisch ausgeschlossen werden. Es muss also eine Möglichkeit geben, wie die Auswirkungen einer fälschlicherweise als Konflikt gewerteten negativen Empfehlung rückgängig gemacht werden.

Diese Überlegungen werden durch die Einführung einer zweiten Ereignisart, der *Rehabilitierung*, berücksichtigt. Eine Einheit erhält dieses Ereignis für ihre Glaubensbildung, wenn sie erkennt, dass ein Konflikt, den sie früher bei Erhalt einer negativen Empfehlung postuliert hat, in Wirklichkeit keiner ist. Das Beispiel aus Abschnitt 7.3.1 zeigt, unter welchen Umständen dieser Fall auftreten kann: Im zweiten Schritt des Beispielablaufs erhält Einheit X die negative Empfehlung der Einheit Y. Zu diesem Zeitpunkt tritt für die Einheit X das Ereignis auf, dass ein Konflikt zwischen Einheit Y und Z wahrgenommen wird. In Folge dessen revidiert Einheit X ihren Typglauben entsprechend. Im dritten Schritt des Ablaufs erhält Einheit X allerdings die Selbstempfehlung der Einheit Z und erkennt, dass die Transaktion zwischen Einheit Y und Z konfliktlos abgelaufen ist. In dieser Situation stellt die Einheit X ihren ursprünglichen Typglauben über die Einheit Z wieder her und rehabilitiert sie dadurch. Wir fassen also zusammen, dass es immer dann zu Rehabilitationen kommt, wenn eine Einheit zunächst negativ empfohlen wird und anschließend durch das Vorzeigen einer entsprechenden Quittung diese Empfehlung widerlegt.

Typbeweis. Bei Erhalt eines Typbeweis gibt es zwei prinzipielle Möglichkeiten. Welche von ihnen zutrifft, hängt davon ab, ob bereits bekannt ist, dass die Einheit, deren Inkonsistenz im Typbeweis bewiesen wird, strategisch ist. Ist dies bereits bekannt, so birgt der Typbeweis keine neue Erkenntnis und wird folglich ignoriert. Andernfalls ist der Glauben über die Einheit, die durch den Typbeweis als inkonsistent handelnd erkannt worden ist, zu revidieren. Diese Einheit ist mit Sicherheit strategisch, da normative Einheiten sich nicht zu inkonsistenten Aussagen festlegen. Daher stellt der Typbeweis eine *Typinformation* dar und der Glaube über die inkonsistent handelnde Einheit wird entsprechend der Vorschriften aus Abschnitt 6.3.3 revidiert.

Selbstempfehlung. Bei Erhalt einer Selbstempfehlung muss eine Einheit überprüfen, ob die in der Empfehlung enthaltenen Quittungen im Widerspruch zu einer zuvor erhaltenen negativen Empfehlung stehen. Ist dies der Fall, so werden zwei Ereignisse der Glaubensrevision ausgelöst, die nacheinander abzuarbeiten sind:

- Der Selbstempfehler ist zu rehabilitieren, da er zuvor zu unrecht negativ empfohlen worden ist.
- Die Einheit, die die inkonsistente negative Empfehlung ausgestellt hat, wird durch die Selbstempfehlung als strategische Einheit erkannt. Dies stellt eine Typinformation über diese Einheit dar, die zu einer entsprechenden Glaubensrevision führt.

Gemäß dieser Darstellung dient das Vorzeigen einer Quittung in einer Selbstempfehlung alleine zur Widerlegung von inkonsistenten negativen Empfehlungen. Kommt es zu keiner solchen Widerlegung, so wird der Glauben über den Selbstempfehler nicht revidiert. Es stellt sich hierbei

die Frage, ob das Vorzeigen von Quittungen nicht immer zu einer Glaubensrevision führen sollte. Genauer gesagt erscheint es auf den ersten Blick plausibel, dass eine Einheit, die ihr kooperatives Transaktionsverhalten durch entsprechende Quittungen belegt, aufgewertet werden soll. Im verbleibenden Teil dieses Abschnitts zeigen wir jedoch, dass eine solche Berücksichtigung von Quittungen nicht sinngemäß ist.

Was würde passieren, wenn das Vorzeigen einer Quittung tatsächlich zu einer Aufwertung des Glaubens über den Selbstempfehler führen würde? In diesem Fall hätten die strategischen Einheiten einen Anreiz, sich gegenseitig eine Reihe von Quittungen auszustellen und sich damit jeweils selbst zu empfehlen. Dies stellt insofern ein Problem dar, als sich Quittungen damit nicht mehr auf Transaktionen beziehen würden, die tatsächlich stattgefunden haben. Die direkte Berücksichtigung von Quittungen für die Glaubensrevision würde daher nur Verschwörungen von strategischen Einheiten fördern, deren unberechtigte Selbstempfehlungen den Prozess der Glaubensbildung beeinträchtigen würde.

Dies zeigt, dass Selbstempfehlungen in der Glaubensbildung anders als negative Empfehlungen zu behandeln sind. Das Vorzeigen von Quittungen in einer Selbstempfehlung darf nur zur Widerlegung von inkonsistenten negativen Empfehlungen führen. Eine darüber hinaus gehende Berücksichtigung ist mit den Beweismittelarten, die wir bisher vorgestellt haben, nicht sinnvoll. In Kapitel 8 werden wir diese Einschränkung für Selbstempfehlungen beseitigen, indem wir soziale Beweismittel einführen. Im Gegensatz zu Quittungen eignen sich diese dazu, Selbstempfehlungen derart zu unterstützen, dass der Glauben über den jeweiligen Selbstempfehler aufgewertet werden kann, ohne Verschwörungen zu fördern. Eine entsprechende Erweiterung des Empfehlungssystems wird in Abschnitt 8.2.2 durchgeführt.

7.4.2 Einbeziehung der Ereignisse in die Glaubensrevision

Durch den Erhalt von Empfehlungen können zwei Arten von Ereignissen ausgelöst werden, die zur Glaubensrevision führen. Es handelt sich bei ihnen um den Konflikt und die Rehabilitierung. Im Folgenden beschäftigen wir uns damit, wie die Glaubensrevision für diese beiden Ereignisarten auf probabilistisch fundierte Weise durchzuführen ist. Dabei greifen wir wie schon in Abschnitt 6.3.3 auf die Odds-Darstellung zurück, um die Revision des Typglaubens übersichtlich darzustellen.

Konflikt. Bei der Beschreibung der Revisionsvorschriften für Konflikte gehen wir davon aus, dass Einheit X einen Konflikt zwischen den Einheiten Y und Z erkannt hat. Dieser bezieht sich auf eine Transaktion des Kontexts γ. Weiterhin nehmen wir an, dass dieser Konflikt aufgrund einer negativen Empfehlung der Einheit Y über Einheit Z wahrgenommen worden ist. Bei dieser Ausgangssituation stellt sich für Einheit X die Frage, wie ihr Typglaube über die Einheiten Y und Z zu revidieren ist. Auf diese Frage gehen wir zunächst für die negativ empfohlene Einheit, nämlich Einheit Z, ein.

Um eine probabilistisch fundierte Revision des Glaubens über Einheit Z durchzuführen, müssen wir auf die Bayes-Formel zurückgreifen: Der Konflikt zwischen den Einheiten Y und Z zeigt an, dass mindestens eine von ihnen betrogen hat. In Anlehnung an Abschnitt 6.3.3 notieren wir diese Aussage mit $D^{(b)}_{YZ}$. Dabei ist für die außenstehende Einheit X nicht erkennbar, welche der beiden Einheiten tatsächlich betrogen hat. Die Formel 6.6 des TIB-Modells erlaubt eine Einschätzung der Wahrscheinlichkeit dafür, dass es zu dem Konflikt kommt. Dabei übernehmen wir den Formalismus des TIB-Modells: Die Terme $p_n(\gamma)$ und $p_u(\gamma)$ beziehen sich weiterhin auf die kontextabhängige Einschätzung der strategischen Einhaltung und des unbeabsichtigten Betrugsverhaltens. Die subjektive Wahrscheinlichkeit der Einheit X darüber, dass Einheit Y in einer

Transaktion mit Kontext γ betrügt, wird mit $p_X(C_Y^{(b)}|\gamma)$ ausgedrückt. Dies führt zur folgenden Einschätzung der Wahrscheinlichkeit eines Konflikts zwischen Einheit Y und Z:

$$\begin{aligned} p_X(D_{YZ}^{(b)}|\gamma) &= 1 - p_X(C_Y^{(b)}|\gamma) \cdot p_X(C_Z^{(b)}|\gamma) \\ &= 1 - p_X(C_Y^{(b)}|\gamma) \cdot [p_X(N_Z) + p_X(S_Z) \cdot p_n(\gamma)] \cdot \overline{p}_u(\gamma) \\ &= 1 - [p_X(N_Y) + p_X(S_Y) \cdot p_n(\gamma)] \cdot [p_X(N_Z) + p_X(S_Z) \cdot p_n(\gamma)] \cdot \overline{p}_u(\gamma)^2 \end{aligned} \quad (7.1)$$

Die Einschätzung der Wahrscheinlichkeit, dass es zu einem Konflikt kommt, hängt also von dem prioren Typglauben über die beiden Einheiten Y und Z ab[2]. Auf der Grundlage dieser Einschätzung lässt sich die Vorschrift für die Revision des Glaubens über die Einheit Z gemäß der Bayes-Formel wie folgt angeben:

$$\begin{aligned} p_X(N_Z|D_{YZ}^{(b)}, \gamma) &= \frac{p_X(D_{YZ}^{(b)}|N_Z, \gamma) \cdot p_X(N_Z)}{p_X(D_{YZ}^{(b)}|\gamma)} \\ \hat{p}_X(N_Z|D_{YZ}^{(b)}, \gamma) &= \hat{p}_X(N_Z) \cdot r_K(Y, \gamma) \\ &= \hat{p}_X(N_Z) \cdot \frac{p_X(D_{YZ}^{(b)}|S_Z, \gamma)}{p_X(D_{YZ}^{(b)}|N_Z, \gamma)} \end{aligned} \quad (7.2)$$

Für die Angabe des Revisionsfaktors r_K der Odds-Darstellung müssen wir zwei bedingte Wahrscheinlichkeiten bestimmen:

$$\begin{aligned} p_X(D_{YZ}^{(b)}|N_Z, \gamma) &= 1 - p_X(C_Y^{(b)}|\gamma) \cdot p_X(C_Z^{(b)}|N_Z, \gamma) \\ &= 1 - p_X(C_Y^{(b)}|\gamma) \cdot \overline{p}_u(\gamma) \\ &= 1 - [p_X(N_Y) + p_X(S_Y) \cdot p_n(\gamma)] \cdot \overline{p}_u(\gamma)^2 \\ p_X(D_{YZ}^{(b)}|S_Z, \gamma) &= 1 - p_X(C_Y^{(b)}|\gamma) \cdot p_X(C_Z^{(b)}|S_Z, \gamma) \\ &= 1 - p_X(C_Y^{(b)}|\gamma) \cdot p_n(\gamma) \cdot \overline{p}_u(\gamma) \\ &= 1 - [p_X(N_Y) + p_X(S_Y) \cdot p_n(\gamma)] \cdot p_n(\gamma) \cdot \overline{p}_u(\gamma)^2 \end{aligned} \quad (7.3)$$

Abschließend erhalten wir also für den Revisionsfaktor r_K:

$$r_K(Y, \gamma) = \frac{p_X(D_{YZ}^{(b)}|S_Z, \gamma)}{p_X(D_{YZ}^{(b)}|N_Z, \gamma)} = \frac{1 - [p_X(N_Y) + p_X(S_Y) \cdot p_n(\gamma)] \cdot p_n(\gamma) \cdot \overline{p}_u(\gamma)^2}{1 - [p_X(N_Y) + p_X(S_Y) \cdot p_n(\gamma)] \cdot \overline{p}_u(\gamma)^2} \quad (7.4)$$

Dadurch, dass sich dieser Quotient nicht weiter vereinfachen lässt, ist diese Formel im Vergleich zu den Revisionsfaktoren r_C und r_D aus Abschnitt 6.3.3 schwieriger zu interpretieren. Bei der Besprechung des Revisionsfaktors gehen wir deshalb schrittweise vor:

- *Strategische Einhaltung:* Der Zähler des Revisionsfaktors r_K unterscheidet sich einzig dadurch von seinem Nenner, dass die Wahrscheinlichkeit strategischer Einhaltung $p_n(\gamma)$ zweimal auftritt. Daraus ergibt sich, dass der Zähler immer größer als der Nenner ist und

[2]Ein weiterer interessanter Aspekt der Einschätzung ist, dass die Wahrscheinlichkeit unbeabsichtigten Verhaltens $p_u(\gamma)$ nur als Quadrat eingeht. Dies ist darin begründet, dass beide Transaktionspartner unbeabsichtigterweise betrügen können. Darin stimmt die Einschätzung mit der Revisionsvorschrift zur Berücksichtigung von Transaktionserfahrungen aus Abschnitt 6.3.3 überein. Jene Vorschriften mussten zu diesem Zweck allerdings die Ebene der Wahrnehmung einführen, da nur so die Möglichkeit eigenen unbeabsichtigten Fehlverhaltens in der Revisionsvorschrift festgehalten werden kann.

der Revisionsfaktor somit größer als eins ist. Damit erhöht sich durch die Revision die Odds-Darstellung $\hat{p}_X(N_Z)$ und die subjektive Wahrscheinlichkeit $p_X(N_Z)$ verringert sich. Anschaulich bedeutet dies, dass eine am Konflikt beteiligte Einheit immer abgewertet wird. Das Ausmaß der Abwertung hängt unter anderem von der Wahrscheinlichkeit strategischer Einhaltung ab. Dies lässt sich an den zwei Extremen der strategischen Einhaltung veranschaulichen:

  - $p_n(\gamma) = 0$: Der Revisionsfaktor vereinfacht sich zu $\frac{1}{1-p_X(N_Y)\cdot\overline{p}_u(\gamma)^2}$. Je mehr die Einheit Y als normativ angesehen wird und je unwahrscheinlicher unbeabsichtigtes Fehlverhalten ist, desto stärker wird also die Einheit Z abgewertet. Dies ist sinnvoll, da es dann wahrscheinlicher ist, dass der Konflikt durch den Betrug der Einheit Z ausgelöst wurde.
  - $p_n(\gamma) = 1$: Der Revisionsfaktor vereinfacht sich zu 1. Dies ergibt sich daraus, dass bei dieser Einschätzung der strategischen Einhaltung strategische Einheiten sich wie normative Einheiten verhalten. Daher lässt das Auftreten eines Konflikts keine Rückschlüsse auf den Typ der beteiligten Einheiten zu.

- *Unbeabsichtigtes Betrugsverhalten:* Auffällig ist, dass die Wahrscheinlichkeit unbeabsichtigten Betrugsverhaltens lediglich in der Form $\overline{p}_u(\gamma)^2$ in die Formel eingeht. Dieser Ausdruck gibt an, mit welcher Wahrscheinlichkeit beide Einheiten Y und Z sich kooperativ verhalten, wenn sie dies auch beabsichtigen. Je geringer diese Wahrscheinlichkeit (etwa aufgrund häufiger Kommunikationsabbrüche) ist, desto weniger stark fällt die Glaubensrevision aus[3]. Dies ist sinnvoll, da dann der Konflikt eine geringere Aussagekraft über die Normativität der beteiligten Einheiten besitzt.

- *Kein unmögliches Ereignis:* Das Auftreten eines Konflikts stellt nur dann ein unmögliches Ereignis dar, wenn sowohl der Zähler als auch der Nenner des Revisionsfaktors null sind. Dies erfordert jedoch $p_n(\gamma) = 1$ und $p_u(\gamma) = 0$. Dies trifft gemäß dem Systemmodell aber nicht zu, da weder strategisches noch unbeabsichtigtes Betrugsverhalten ausgeschlossen werden kann. Somit benötigen wir keine gesonderten Revisionsvorschriften für das Auftreten unmöglicher Ereignisse.

- *Einfluss des Typglaubens:* Der Revisionsfaktor r_K zeichnet sich dadurch aus, dass in seine Berechnung der Typglauben über die Einheit Y eingeht. Dies schlägt sich unter anderem auch darin nieder, dass der Revisionsfaktor nicht nur eine Funktion des Kontexts γ sondern auch des jeweiligen Transaktionspartners Y ist. Je normativer diese Einheit Y erscheint, desto eher wird ihr zugetraut, dass sie kooperatives Verhalten beabsichtigt. Dies wird in der Formel durch den Term $[p_X(N_Y) + p_X(S_Y) \cdot p_n(\gamma)]$ festgehalten, der gemäß dem TIB-Modell eben diese Absicht kooperativen Verhaltens einschätzt. Je größer dieser Term wird, desto größer fällt der Revisionsfaktor r_K aus[4]. Wir erhalten also, dass der Empfohlene Z umso stärker abgewertet wird, je normativer der Aussteller Y der negativen Empfehlung erscheint.

Die Besprechung zeigt also, dass erkannte Konflikte durch die vorgestellte Revisionsvorschrift

[3] Durch eine Verringerung von $\overline{p}_u(\gamma)^2$ wird auch der Subtrahend des Zählers und Nenners kleiner, so dass sich der Revisionsfaktor in die Richtung der Zahl 1 bewegt.

[4] Dies ergibt sich daraus, dass der Nenner der Formel sich stärker vermindert als der Zähler, bei dem der zusätzliche Faktor $p_n(\gamma)$ für eine Abschwächung dieser Änderung sorgt.

auf probabilistisch fundierte Weise berücksichtigt werden. Außerdem wird deutlich, dass die Entwurfsprinzipien des Systementwurfs umgesetzt werden:

- *Erstes Entwurfsprinzip:* Für eine Einheit führt die Beteiligung an einem Konflikt zu einer umso größeren Abwertung, je normativer ihr Transaktionspartner den anderen Einheiten erscheint. Um eine starke Abwertung zu vermeiden, wird eine betrugswillige strategische Einheit daher bevorzugt in Transaktionen mit anderen strategischen Einheiten betrügen. Damit verhält sich jede strategische Einheit normativen Einheiten gegenüber kooperativer als strategischen Einheiten gegenüber. Dies trägt zur Paradoxie strategischen Verhaltens bei.

- *Zweites Entwurfsprinzip:* Je normativer eine Einheit angesehen wird, desto stärker ist das Gewicht ihrer negativen Empfehlungen bei der Glaubensrevision anderer Einheiten. In der Analyse des Abschnitts 7.6.2 werden wir erkennen, dass dieser Zusammenhang einen weiteren Vorteil dafür schafft, als normativ angesehen zu werden.

- *Drittes Entwurfsprinzip:* Um als normativ angesehen zu werden, müssen strategische Einheiten vermeiden, an Konflikten beteiligt zu sein. Hierfür ist wiederum erforderlich, dass sie von Betrugsverhalten absehen. Je unwahrscheinlicher unbeabsichtigtes Betrugsverhalten ist, desto schneller erscheint nämlich eine im Konflikt beteiligte Einheit als strategisch.

In der bisherigen Darstellung haben wir die Frage erörtert, wie der Typglauben über die negativ empfohlene Einheit Z zu revidieren ist. Analoge Überlegungen könnten wir auch für die Einheit Y anstellen, da der Konflikt ebenso von ihr verursacht worden sein könnte. Als Folge davon käme es zu einer Abwertung der Einheit Y. Dies würde allerdings das Problem hervorrufen, dass das Ausstellen von negativen Empfehlungen einen erheblichen Nachteil mit sich bringt. Somit würde es unter Umständen erst gar nicht zum Ausstellen negativer Empfehlungen kommen. Um die Verfügbarkeit negativer Empfehlungen nicht zu gefährden, müssen wir demnach bei der Definition der Vorschriften darauf verzichten, dass der Glauben über den Aussteller der negativen Empfehlung revidiert wird.

Dass die einseitige Bewertung von Konflikten dennoch keine Einschränkung darstellt, zeigt die folgende Überlegung: Nicht nur Einheit Y ist in der Lage, über ihren Konflikt mit der Einheit Z durch das Ausstellen einer entsprechenden negativen Empfehlung zu berichten. Umgekehrt ist es auch der Einheit Z möglich, die Einheit Y negativ zu empfehlen. Das Ausstellen einer solchen negativen Empfehlung führt dazu, dass auch Einheit Y entsprechend der obigen Revisionsvorschriften abgewertet wird. Dadurch erhalten wir, dass der Glauben über die beiden am Konflikt beteiligten Einheiten revidiert wird. Dass es auch tatsächlich zum beidseitigen Ausstellen von negativen Empfehlungen kommt, wird die Analyse des Abschnitts 7.6.2 zeigen.

Rehabilitierung. Bei dem Ereignis der Rehabilitierung gehen wir von folgender Konstellation aus: Einheit X hat zuvor von der Einheit Y eine negative Empfehlung über die Einheit Z erhalten und diese entsprechend abgewertet. Zu einem späteren Zeitpunkt widerlegt die Einheit Z jedoch diese negative Empfehlung. Diese Konstellation stimmt mit der aus Abbildung 7.6 überein. In diesem Fall muss es zur Rehabilitierung der Einheit Z kommen. Genauer gesagt muss Einheit X die Auswirkungen der vorigen Abwertung rückgängig gemacht werden. Um dieses Ziel zu erreichen, sind zwei Methoden denkbar:

- *Exakte Rehabilitierung:* Bei jeder Berücksichtigung eines Konflikts merkt sich eine Einheit, mit welchem Revisionsfaktor sie die negativ empfohlene Einheit abgewertet hat. In der

oben genannten Konstellation muss sich Einheit X demnach den Revisionsfaktor $r_K(Y, \gamma)$ merken. Kommt es anschließend zur Rehabilitierung der Einheit Z, so ist der Typglaube in seiner Odds-Darstellung durch diesen Revisionsfaktor zu dividieren. Wir erhalten also als posterioren Typglauben bei Eintreten der Rehabilitierung R_{YZ} Folgendes:

$$\hat{p}_X(N_Z|R_{YZ}, \gamma) = \hat{p}_X(N_Z) \cdot \frac{1}{r_K(Y, \gamma)} = \hat{p}_X(N_Z) \cdot r_R(Y, \gamma) \tag{7.5}$$

Dabei gibt $r_R(Y, \gamma)$ den Revisionsfaktor der Rehabilitierung an. Er ist der Kehrwert des Revisionsfaktors des Konflikts.

- *Speicherschonende Rehabilitierung:* Das Problem der Methode der exakten Rehabilitierung liegt darin, dass für jede Berücksichtigung von Konflikten der dabei angewendete Revisionsfaktor gespeichert werden muss. Dies führt zu einem erheblichen Speicher- und Verwaltungsaufwand. Es erscheint daher wünschenswert, dass der Revisionsfaktor nicht gespeichert sondern im Nachhinein rekonstruiert wird. Auf den ersten Blick ist dies problemlos möglich, da die Größe des Revisionsfaktors r_K lediglich von dem negativen Empfehler Y und dem Transaktionskontext γ abhängt. Beides lässt sich aus der Quittung, deren Erhalt zur Rehabilitierung der Einheit Z führt, ablesen. Allerdings könnte sich der Typglaube über die Einheit Y in der Zwischenzeit geändert haben. Wird bei der Rekonstruktion des Revisionsfaktors $r_K(Y, \gamma)$ der aktuelle Typglaube über Einheit Y eingesetzt, so kann dies also zu einer Über- oder Unterschätzung der Revisionsfaktors führen. Hierbei gilt, dass der ursprünglich eingesetzte Revisionsfaktor umso genauer rekonstruiert werden kann, je kürzer die Zeit zwischen Abwertung und Rehabilitierung der Einheit Z ist. In diesem Fall ist eine zwischenzeitliche Revision des Glaubens über die Einheit Y unwahrscheinlicher. Nichtsdestoweniger kann diese zweite speicherschonende Methode zu Ungenauigkeiten bei der Rehabilitierung einer Einheit führen.

Bei der Wahl der anzuwendenden Methode muss jede Einheit für sich selbst abwägen, ob die Exaktheit der ersten Methode den damit einhergehenden Speicheraufwand rechtfertigt oder nicht. Wie bereits bei der Besprechung der Vorschriften für das Empfehlungsverhalten in Abschnitt 7.3.2 ist es für unsere weiteren Betrachtungen nicht von Belang, wie die Wahl der einzelnen Einheiten jeweils ausfällt. Wichtig ist lediglich, dass es zu einer entsprechenden Rehabilitierung kommt, sobald die Inkonsistenz des negativen Empfehlers erkannt wird.

7.5 Beweismittel- und Wissensverwaltung

Durch die Teilnahme an Transaktionen und den Erhalt von Empfehlungen erhält eine Einheit Beweismittel. Diese muss die Einheit bei sich lokal ablegen, da sie für das Ausstellen von Empfehlungen und das Ziehen von Schlussfolgerungen für die Glaubensrevision benötigt werden. Es bietet sich daher an, die Wissensverwaltung einer Einheit zu einer kombinierten Beweismittel- und Wissensverwaltung auszubauen.

Mit der Umsetzung dieses Zieles beschäftigen wir uns in diesem Abschnitt. Zunächst identifizieren wir in Abschnitt 7.5.1 die Anforderungen an diese kombinierte Beweismittel- und Wissensverwaltung. Anschließend wird in Abschnitt 7.5.2 die Umsetzung dieser Anforderungen besprochen.

Tabelle 7.1: Beweismittel und Wissen für die Ablage in der Verwaltungskomponente

Art der Information	**Kategorie**	**Erhalt in**	**Struktur**
Vertrag	Beweismittel	Transaktion, negative Empfehlung, Typbeweis	TID, Aussteller, Transaktionspartner, Aktionsbeschreibung, Verfalldatum
Quittung	Beweismittel	Transaktion, Selbstempfehlung	TID, Aussteller, Transaktionspartner, Aktionsbeschreibung, Verfalldatum
Kooperation	Wissen	Transaktion	TID,Transaktionspartner
Betrug	Wissen	Transaktion	TID,Transaktionspartner
Negative Empfehlung	Beweismittel	negative Empfehlung, Typbeweis	TID,Aussteller,Transaktionspartner

7.5.1 Anforderungen

Im Folgenden untersuchen wir, welchen Anforderungen die Beweismittel- und Wissensverwaltung gerecht werden muss. Die Anforderungen leiten sich direkt aus den Bedürfnissen des Transaktionsprotokolls, des Empfehlungssystems und der Glaubensbildung ab: **(1)** Beweismittel und Wissen, das in Transaktionen oder durch Empfehlungen erlangt worden ist, muss die Verwaltungskomponente ablegen können. **(2)** Für das Ausstellen von Empfehlungen und die Glaubensrevision ist ein assoziativer Zugriff auf die abgelegten Beweismittel und das Wissen notwendig. Da das Empfehlungssystem und die Glaubensbildung an bestimmten Arten von Wissen interessiert ist, das so nicht abgelegt worden ist, muss die Verwaltungskomponente in der Lage sein, aus der Ablage der Beweismittel und des Wissens weiteres Wissen abzuleiten. **(3)** Die in den ersten zwei Punkten geforderte Funktionalität der Verwaltungskomponente darf nicht zu einem unkontrollierten Speicherverbrauch führen, der das Gerät, auf dem sich die jeweilige Einheit befindet, überfordert.

Ablegen von Beweismitteln und Wissen. Im Laufe ihrer Transaktionen erhält eine Einheit einerseits Verträge und Quittungen und andererseits das Wissen über das Verhalten des jeweiligen Transaktionspartners. Um diese Informationen dem Empfehlungssystem und der Glaubensbildung zur Verfügung zu stellen, muss es zu einer Ablage dieser transaktionalen Beweismittel und des Wissens kommen. Ein ähnliches Bild ergibt sich im Empfehlungssystem. Bei Erhalt einer Empfehlung müssen die in ihr mitgelieferten Beweismittel abgelegt werden. Da negative Empfehlungen selbst Beweismittel sind, sind auch sie abzulegen.

Tabelle 7.1 gibt eine Übersicht der Beweismittel und des Wissens, das die Verwaltungskomponente entgegennehmen muss. Verträge und Quittungen erhält eine Einheit nicht nur in ihren Transaktionen, sondern auch in den eingehenden Empfehlungen. Verträge sind in negativen Empfehlungen und Typbeweisen enthalten, während Quittungen im Zuge von Selbstempfehlungen erhalten werden. Negative Empfehlungen als Beweismittel sind Teil von Typbeweisen und sind daher in diesen enthalten.

Bei den abgelegten Informationen handelt es sich um Tupel, die jeweils die entsprechenden Informationen tragen. Die Struktur dieser Tupel ist in Abhängigkeit der Art der Information in der Tabelle eingetragen. Sie erklärt sich für Verträge und Quittungen direkt aus der Beschreibung des Abschnitts 7.2.1. Wissen über Kooperation und Betrug bezieht sich in seinem Transaktions-Identifikator (TID) auf die jeweilige Quittung beziehungsweise Vertrag des Transaktionspartners. Eine negative Empfehlung stellt eine Art negativer Quittung dar. Ihr Aufbau leitet sich daher

von dem einer Quittung ab. Der einzige Unterschied besteht darin, dass die Beschreibung der Aktion und der Zeitraum der Gültigkeit weggelassen werden, da sie bereits im Vertrag enthalten sind, der eine negative Empfehlung zwingendermaßen begleitet.

Zugriff auf Beweismittel und Wissen. Die Ablage von Beweismitteln und Wissen dient dazu, dem Empfehlungssystem und der Glaubensbildung die von ihr benötigten Informationen bereitzustellen. Dabei werden nicht nur diejenigen Informationen benötigt, die direkt abgelegt worden sind. Hinzu kommt eine Art von Wissen, das aus den abgelegten Informationen abgeleitet werden muss, nämlich die *Inkonsistenz* der Festlegungen einer Einheit: Bei einer solchen Inkonsistenz handelt es sich um den Fall, dass eine Einheit bezüglich derselben Transaktion eine Quittung und eine negativen Empfehlung ausgestellt hat. Da aus dieser Information geschlossen werden kann, dass es sich bei der Einheit um eine strategische handelt, geht diese Art der Information als Typinformation in die Glaubensbildung ein. Außerdem zeigt sie an, dass anderen Einheiten gegenüber ein entsprechender Typbeweis geführt werden kann.

Die Verwaltungskomponente hat dafür zu sorgen, dass der Zugriff auf benötigte Informationen assoziativ erfolgen kann. Dies bedeutet, dass die Glaubensbildung und das Empfehlungssystem spezifizieren können, an welcher Art der Informationen sie interessiert sind. Werden solche Informationen verfügbar, so werden sie an die Glaubensbildung oder dem Empfehlungssystem zugestellt.

Eine weitere Anforderung an den Zugriff auf Wissen besteht darin, dass es nachvollziehbar sein soll, aus welchen Beweismitteln es abgeleitet worden ist. Diese Beweismittel stellen den Ausgangspunkt für den Beweis des abgeleiteten Wissens dar. Zum Beispiel muss dem Empfehlungssystem für das Ausstellen eines Typbeweises ersichtlich sein, aus welchem Paar von Quittung und negativer Empfehlung die Inkonsistenz einer Einheit folgt.

Eingeschränkter Speicherverbrauch. Da eine Einheit im Laufe ihrer Teilnahme am Informationssystem ständig Beweismittel und Wissen erlangt und ablegt, wächst die zur Verwaltung notwendige Speichergröße auf unbeschränkte Weise. Dies ist insbesondere für Informationsgeräte wie PDAs kritisch, die über wenige Ressourcen verfügen.

Die Verwaltungskomponente muss also dafür sorgen, dass der Speicherverbrauch sich beschränken lässt. Hierfür ist es unter Umständen notwendig, Beweismittel und Wissen schon vor ihrem Verfalldatum aus der Ablage zu entfernen. Dies bringt allerdings ein Problem hervor: Wird ein Beweismittel im Zuge einer Empfehlung erhalten, das sich zuvor bereits in der Ablage befand, aber mittlerweile entfernt worden ist, so darf es nicht erneut Eingang in die Ablage finden. Ansonsten würde der Sachverhalt, der im Beweismittel dargestellt ist, von der Glaubensbildung erneut berücksichtigt. Die Verwaltungskomponente muss also dafür sorgen, dass aus Speichergründen entfernte Beweismittel nicht mehr abgelegt werden können.

7.5.2 Umsetzung

Im Folgenden diskutieren wir, wie die Beweismittel- und Wissensverwaltung die identifizierten Anforderungen erfüllen kann. Hierfür wird zunächst die Architektur der Verwaltungskomponente vorgestellt. Anschließend gehen wir auf die beiden funktionalen Blöcke der Architektur ein. Dabei handelt es sich um die Ablageverwaltung und die Steuereinheit.

Architektur. Die Anforderungen lassen sich in zwei Bereiche unterteilen. Erstens müssen Beweismittel und Wissen abgelegt und abgerufen werden können. Hierfür ist die Ablageverwaltung

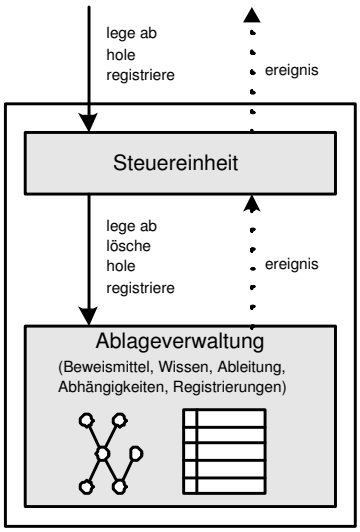

Abbildung 7.7: Architektur der Beweismittel- und Wissensverwaltung

verantwortlich. Zweitens ist der Speicherverbrauch dieser Ablageverwaltung durch eine zusätzliche Komponente, die Steuereinheit, einzuschränken. Diese steht zwischen der Ablageverwaltung einerseits und dem darauf zugreifenden Transaktionsprotokoll, Empfehlungssystem und der Glaubensbildung andererseits. Dadurch ist die Steuereinheit in der Lage, Entscheidungen über das Löschen von bereits abgelegten Beweismitteln oder Wissen zu treffen und die damit verbundenen Probleme zu behandeln.

Abbildung 7.7 zeigt eine Übersicht dieser Architektur. Die Ablageverwaltung ist dafür zuständig, abgelegte Beweismittel und Wissen zu speichern, aus ihnen weiteres Wissen abzuleiten und den assoziativen Zugriff zu ermöglichen. Die einzelnen Arten von Wissen und Beweismitteln, die dabei zum Einsatz kommen, sind in Abschnitt A.1.1 des Anhangs besprochen.

Für den assoziativen Zugriff gibt es zwei Varianten: Wenn im Empfehlungssystem auf Anfrage eine Empfehlung zusammenzustellen ist, wird direkt das Wissen und die Beweismittel, die für die Empfehlung laut Empfehlungsmodell des Abschnitts 7.3.1 von Bedeutung sind, mit dem Primitiv *hole* abgefragt. Die Glaubensbildung hingegen benötigt einen Mechanismus, mit dem sie über neu entstandenes Wissen informiert wird. Hierfür ist initial eine *Registrierung* notwendig, die die Art der benötigten Information der Ablageverwaltung gegenüber beschreibt. Passt neu entstandenes Wissen zu einer Registrierung, so wird dies in einem entsprechenden *Ereignis* der Glaubensbildung gemeldet. Um zu Glaubensrevisionen entsprechend Abschnitt 6.3.3 und 7.4.2 kommen zu können, registriert sich die Komponente der Glaubensbildung somit für eigene Transaktionserfahrungen, neu erhaltene negative Empfehlungen und Typbeweise. Abschnitt A.1.1 des Anhangs gibt die Einzelheiten hierzu.

Die Aufrufe der Ablageverwaltung werden stets durch die Steuereinheit geleitet. Dadurch ist sie in der Lage, Entscheidungen über das vorzeitige *Löschen* von Beweismitteln und Wissen zu fällen. Eine solche Entscheidung ist dahingehend zu optimieren, dass diejenige Information gelöscht wird, die mit der geringsten Wahrscheinlichkeit in Zukunft benötigt wird.

Ablageverwaltung. Eine Implementierung der Ablageverwaltung wird in [Dör04] beschrieben. Sie baut auf das regelbasierte System *Jess* auf [FH03], um die Ableitung von Wissen zu gewährleisten. Die Ableitungsregeln sind in Abschnitt A.1.1 des Anhangs aufgelistet. Durch den Einsatz

des regelbasierten Systems beschränken sich die Aufgaben der Implementierung auf die folgenden Punkte: **(1)** Zwar finden die Aussagen der Beweismittel Eingang in das regelbasierte System. Die zu den Beweismitteln zugehörigen kryptographischen Marken müssen jedoch separat abgelegt werden und dem daraus abgeleiteten Wissen zugeordnet werden. **(2)** Eine Implementierung des Registrierungsmechanismus ist erforderlich, da er nicht vom regelbasierten System zur Verfügung gestellt wird. **(3)** Der Zugriff auf abgelegte Beweismittel und Wissen erfolgt über Schablonen (engl.: templates), die in entsprechende Anfragen an das regelbasierte System umgesetzt werden.

Die Evaluation der Implementierung zeigt, dass der Aufwand für die Ablageverwaltung sehr gering ist[5]. Dies ist insofern bemerkenswert, als die Effizienz der Ablageverwaltung weiter gesteigert werden kann, wenn Techniken der Indexierung angewandt werden [OG02, Obr05].

Steuereinheit. Die Steuereinheit registriert für jede abgelegte Information, wann auf sie zum letzten Mal zugegriffen wurde. Stößt die Ablage an die Obergrenze ihres Speicherverbrauchs, so wird diejenige Information gelöscht, die seit der längsten Zeit nicht mehr verwendet wurde. Diese Verdrängungsstrategie wird in der Literatur *least recently used* (LRU) genannt. Dabei ist zu berücksichtigen, dass es auch indirekte Zugriffe gibt: Auf ein Beweismittel oder auf Wissen wird indirekt zugegriffen, wenn Wissen, das daraus abgeleitet ist, von dem Empfehlungssystem oder der Glaubensbildung angefordert wird.

Das Problem der erneuten Berücksichtigung eines bereits gelöschten Beweismittels wird folgendermaßen gelöst: Wird ein Beweismittel vor dem Ablauf seines Verfalldatums verdrängt, so merkt sich die Steuereinheit, eine eindeutige Kennung des Beweismittels. Bei Verträgen, Quittungen und negativen Empfehlungen besteht sie aus der Kombination aus Identifikator der Transaktion und der Identität des Ausstellers und seines Transaktionspartners. Diese Kennung ist deutlich kürzer als das Beweismittel selbst, da sie nicht alle Bestandteile (insbesondere nicht die Signatur) des Beweismittels mit sich trägt. Das erneute Hinzufügen eines Beweismittels wird also dadurch abgefangen, dass seine Kennung mit der Kennung bereits gelöschter Beweismittel verglichen wird. Nach dem Verfalldatum eines Beweismittels lässt sich auch die Kennung endgültig löschen, da das Beweismittel bei einem erneuten Erhalt aufgrund seines Verfalldatums ignoriert wird. Daher lässt sich der Speicheraufwand nachhaltig begrenzen, ohne dass das Problem einer doppelten Berücksichtigung eines Beweismittels zum Tragen kommt.

7.6 Analyse strategischen Verhaltens

In den vorangehenden Abschnitten sind Normen und Vorschriften für den Umgang mit transaktionalen Beweismitteln aufgestellt worden. Sie erlauben eine verstärkte soziale Kontrolle zwischen den Einheiten des Informationssystems. Aus der Sicht des Systementwurfs ist es erforderlich, dass die Kosten für diese soziale Kontrolle nicht nur von den normativen sondern auch von den strategischen Einheiten getragen werden. Hierfür ist zu zeigen, dass die aufgestellten Normen und Vorschriften von den strategischen Einheiten aus ihrem eigenen Interesse eingehalten werden und somit selbstdurchsetzend sind.

Mit dieser Aufgabe befasst sich dieser Abschnitt. Um strategisches Verhalten vorherzusagen, analysieren wir, welche Verhaltensweisen rational und daher von den strategischen Einheiten zu erwarten sind. Mit dem transaktionalen Verhalten beschäftigen wir uns in Abschnitt 7.6.1, mit

[5]Gemäß der Evaluation in [Dör04] benötigt der assoziative Zugriff weit unter einer Millisekunde, solange sich in der Ablage weniger als 10000 Beweismittel befinden. Über eine solche Größe wächst der Ablage nicht an, wenn die Steuereinheit entsprechend eingestellt ist.

dem Empfehlungsverhalten in Abschnitt 7.6.2. Die Ergebnisse der Analyse werden abschließend in Abschnitt 7.6.3 in Beziehung gebracht. Dies ermöglicht eine Bewertung der Normen und Vorschriften bezüglich ihrer Selbstdurchsetzung.

7.6.1 Transaktionsverhalten

Das Sechs-Wege Transaktionsprotokoll schreibt vor, dass beide Transaktionspartner sich gegenseitig je einen Vertrag und eine Quittung ausstellen. Es ist also im Folgenden zu klären, ob strategische Einheiten sich auch dazu entscheiden, so wie normative Einheiten diese transaktionalen Beweismittel ihrem jeweiligen Transaktionspartner bereitzustellen.

Die zweite Fragestellung, die im Folgenden angegangen wird, beschäftigt sich damit, wie die Existenz transaktionaler Beweismittel strategische Einheiten in ihrem Betrugsverhalten beeinflusst. Die Analyse hierzu zeigt die Paradoxie strategischen Verhalten, wie sie im ersten Entwurfsprinzip des Systementwurfs gefordert wird.

Ausstellung von Verträgen. In Abschnitt 7.2.1 sind wir bereits darauf eingegangen, welche Art der Signalisierung das Ausstellen von Verträgen darstellt. Die dortige Ausführung ergänzen wir im Folgenden durch die Einsichten, die wir beim Entwurf des Empfehlungssystems und der Glaubensbildung erhalten haben.

Im Bezug auf eine Transaktion lassen sich zwei Arten von Einheiten unterscheiden: Eine Einheit kann die Absicht haben, sich in der Transaktion kooperativ zu verhalten. Alternativ dazu kann sie Betrugsverhalten beabsichtigen. Aus der Sicht des Systementwurfs soll das Ausstellen eines Vertrages ein Signal dafür sein, von welcher Art eine Einheit ist. Dazu sind die beiden Bedingungen der glaubhaften Signalisierung zu überprüfen:

- *Kann-Nicht-Bedingung:* Die Bedingung fordert, dass eine Einheit bei der Absicht zu Betrug keinen Vertrag ausstellt. Die Diskussion in Abschnitt 7.2.1 hat zwar gezeigt, dass diese Bedingung nicht unter allen Umständen erfüllt ist. Die Möglichkeit, negativ empfohlen zu werden, ist aber dafür verantwortlich, dass der Betrugsvorteil stark vermindert, wenn nicht sogar eliminiert wird. Dies gilt umso mehr, als der Transaktionspartner normativ erscheint, da dann seine negative Empfehlung ein stärkeres Gewicht findet.

- *Kann-Bedingung:* Hierbei ist die Frage zu klären, ob eine strategische Einheit mit Absicht zur Kooperation einen Vertrag ausstellt oder nicht. Ist ihr Transaktionspartner normativ, so ist das Ausstellen des Vertrags gemäß dem Sechs-Wege Transaktionsprotokoll eine Voraussetzung für das Fortführen der Transaktion. Die Frage stellt sich also nur dann, wenn sich die strategische Einheit sicher ist, dass ihr Transaktionspartner strategisch ist. Allerdings wird sie auch dann den Vertrag ausstellen, um ihm gegenüber normativ zu erscheinen. Dies ergibt sich direkt aus der Paradoxie strategischen Verhaltens: Strategische Einheiten betrügen sich bevorzugt gegenseitig, da sie bei Betrug eines normativ erscheinenden Transaktionspartners höhere Folgekosten antizipieren.

Wie wirkt sich die Sequentialität der Vertragsausstellung, die im Sechs-Wege Protokoll vorgesehen ist, auf diese Betrachtung aus? Die Einheit in der sicheren Position ist in der Lage, nach Erhalt des Vertrags ihren Transaktionspartner zu betrügen, indem sie selbst keinen Vertrag ausgestellt. Es gibt allerdings zwei Gründe, warum solches Betrugsverhalten auch von strategischen Einheiten nicht zu erwarten ist: **(1)** Der Erhalt des Vertrags des Transaktionspartners ermöglicht zwar das Ausstellen einer negativen Empfehlung über ihn. Dies bringt aber keinen Vorteil mit

sich, da die Revisionsvorschriften der Glaubensbildung das negative Empfehlen nicht belohnen. **(2)** Durch den Betrug ist der Transaktionspartner weniger zu zukünftigen Transaktionen bereit. Somit verursacht das Vorenthalten des Vertrags Folgekosten. Insgesamt erhalten wir, dass diese Art von Betrug also keinen Vorteil aber Kosten mit sich bringt.

Ausstellung von Quittungen. Durch das Ausstellen einer Quittung signalisiert eine Einheit, dass sie ihren Transaktionspartner nicht negativ empfehlen wird. Abschnitt 7.2.1 hat die Gültigkeit der Kann-Nicht-Bedingung gezeigt: Eine Einheit, die negativ empfehlen will, stellt keine Quittung aus, um sich nicht der Gefahr auszusetzen, von den anderen Einheiten als inkonsistent handelnd und damit strategisch seiend erkannt zu werden.

Dass auch die Kann-Bedingung erfüllt ist, zeigt die folgende Überlegung: Für eine Einheit, die kein negatives Empfehlen beabsichtigt, bringt das Ausstellen der Quittung (abgesehen vom Ausführen der dazu notwendigen kryptographischen Operation) keinen Nachteil mit sich. Sie gibt dadurch lediglich die Möglichkeit zum negativen Empfehlen auf, die sie sowieso nicht wahrnehmen wollte. Auf der anderen Seite hat das Ausstellen der Quittung zum Vorteil, dass nur so der Transaktionspartner das eigene Verhalten als kooperativ ansieht und dadurch eher zu zukünftigen Transaktionen bereit ist. Die Erfüllung der Kann-Bedingung ergibt sich also aus Gründen, die analog zu denen für das Ausstellen von Verträgen sind.

Betrugsverhalten. In Abschnitt 7.4.2 ist bereits angeklungen, dass der Einsatz von transaktionalen Beweismitteln die vom ersten Entwurfsprinzip geforderte Paradoxie strategischen Verhaltens weiter verstärkt. Im Folgenden gehen wir genauer auf diesen Punkt ein.

Das transaktionale Verhalten einer Einheit kann in zwei verschiedenen Kategorien aufgeteilt werden:

- *Wahl des Transaktionspartners:* Jede Einheit muss sich entscheiden, mit welchen Einheiten sie Transaktionen eingehen will. Hierbei gilt sowohl für normative als auch strategische Einheiten, dass sie Transaktionen mit normativ erscheinenden Einheiten bevorzugen. Dies ergibt sich daraus, dass normative Einheiten nie Betrugsverhalten beabsichtigen. Um die Wahrscheinlichkeit für einen Konflikt mit dem eigenen Transaktionspartner zu minimieren, werden also bevorzugt normativ erscheinende Einheiten als Transaktionspartner angenommen.

- *Wahl der Verhaltens in der Transaktion:* Befindet sich eine strategische Einheit in einer Transaktion mit einem ihr normativ erscheinenden Transaktionspartner, so wird sie vermeiden, ihn zu betrügen. Andernfalls würde es zu einem Konflikt kommen, durch den es zum Ausstellen negativer Empfehlungen käme. Dadurch, dass der Transaktionspartner normativ erscheint, würde die Abwertung in Folge einer solchen negativen Empfehlung besonders stark ausfallen. Im Gegensatz dazu sind die Folgekosten für das Betrügen von strategisch erscheinenden Einheiten weitaus geringer. Sie können sich im Rahmen des Empfehlungssystems schlechter verteidigen, wenn es aufgrund des Betrugsverhaltens zum Konflikt kommt. Dies liegt am geringen Gewicht, das ihren Empfehlungen beigemessen wird. Ein weiterer Grund dafür, bevorzugt strategische Einheiten zu betrügen, liegt darin, dass diese einen in zukünftigen Transaktionen eventuell selbst betrügen werden. Der Wegfall von Gelegenheiten zu zukünftigen Transaktionen mit ihnen führt also zu einem geringeren Verlust und kann eher hingenommen werden.

In den beiden Kategorien transaktionalen Verhaltens werden normative Einheiten bevorzugt von den strategischen Einheiten behandelt. Daraus ergibt sich die Paradoxie strategischen Verhaltens und das erste Entwurfsprinzip wird eingehalten. Als direkte Folge davon wünscht jede strategische Einheit, anderen Einheiten gegenüber als normativ zu erscheinen, um in den Genuss dieser bevorzugten Behandlung zu kommen. Damit ist auch das zweite Entwurfsprinzip erfüllt. Im nachfolgenden Abschnitt befassen wir uns mit dem Empfehlungsverhalten der strategischen Einheiten. Als Ergebnis der dortigen Analyse erhalten wir, dass nicht nur normative sondern auch strategische Einheiten ihre Transaktionserfahrungen mitteilen. Die dadurch erreichte Wahrnehmbarkeit Norm-bezogenen Verhaltens führt zur Einhaltung des dritten Entwurfsprinzips. Damit entsprechen die Vorschriften des Systementwurfs zum Einsatz der transaktionalen Beweismitteln den Vorgaben aus Kapitel 5.

7.6.2 Empfehlungsverhalten

In diesem Abschnitt wenden wir uns der Fragestellung zu, ob strategische Einheiten dazu bereit sind, Empfehlungen auszustellen. Diese aktive Beteiligung am Empfehlungssystem ist erforderlich, weil nur dann die Transaktionserfahrungen und das Wissen der strategischen Einheiten anderen Einheiten zugänglich wird.

Bei Selbstempfehlungen und Typbeweisen stellt sich die Situation wie folgt dar: Es ist im Interesse einer strategischen Einheit, sich selbst zu empfehlen. Nur so ist sie in der Lage, sich vor Diffamierungen anderer Einheiten zu schützen. Ein ähnliches Bild ergibt sich auch bei Typbeweisen: Ist eine strategische Einheit in einen Konflikt mit einer anderen Einheit verwickelt, so ist es in ihrem Interesse, ihren Transaktionspartner bei den anderen Einheiten zu diskreditieren. Nur dadurch kann die Glaubwürdigkeit einer etwaigen negativen Empfehlung des Transaktionspartners vermindert und die damit verbundene eigene Abwertung abgeschwächt werden. Wenn ein Typbeweis über den Transaktionspartner der strategischen Einheit verfügbar ist, so ist die Weitergabe dieses Typbeweises in ihrem Interesse. Dies liegt daran, dass der Transaktionspartner durch den Typbeweis am effektivsten diskreditiert wird. Zusammenfassend erhalten wir also, dass es einen inhärenten Anreiz zum Ausstellen von Selbstempfehlungen und Typbeweisen gibt. Daher stellen auch strategische Einheiten solche Empfehlungen aus.

Bei negativen Empfehlungen gibt es keinen solchen inhärenten Anreiz. Durch die Vorschriften der Glaubensrevision sorgen wir zwar dafür, dass der Aussteller einer negativen Empfehlung aufgrund seiner Empfehlung nicht abgewertet wird. Es ist aber unklar, welchen Vorteil sich eine Einheit durch das negative Empfehlen verschafft. So ist zum Beispiel im Empfehlungssystem keine Belohnung für das Ausstellen von negativen Empfehlungen vorgesehen.

Im Folgenden untersuchen wir daher, ob und unter welchen Umständen strategische Einheiten negativ empfehlen. Dabei gehen wir von der Ausgangssituation aus, dass eine strategische Einheit Y in einen Konflikt verwickelt ist. Für sie stellt sich die Frage, ob sie ihren Transaktionspartner Z einer anderen Einheit X gegenüber negativ empfehlen soll. Die nachfolgende Analyse dieser Situation untersucht zunächst, in welcher Hinsicht das Ausstellen einer solchen negativen Empfehlung für die Einheit Y von Vorteil ist. Dies ermöglicht die Modellierung der Ausgangssituation als ein Spiel. Die Analyse dieses Spiels bietet einen Aufschluss darüber, ob das Ausstellen negativer Empfehlungen rational ist. Die Schlussfolgerungen, die aus der Analyse gezogen werden, sind in einem abschließenden Fazit zusammengestellt.

Vorteil des negativen Empfehlens. Wie bereits erwähnt, bringt das Ausstellen einer negativen Empfehlung keinen direkten Vorteil mit sich. Der Adressat der Empfehlung revidiert gemäß

der Revisionsvorschriften lediglich seinen Glauben über den negativ Empfohlenen. Auf den ersten Blick erscheint es daher so, als ob kein Anreiz für das Ausstellen negativer Empfehlungen existiert. Im Folgenden zeigen wir, dass dem jedoch nicht so ist.

Der entscheidende Punkt unserer Betrachtung ist, dass nicht nur die Einheit Y sondern auch die Einheit Z negativ empfehlen kann. Wenn diese Einheit Z tatsächlich eine negative Empfehlung an die Einheit X ausstellt, sind drei Fälle zu unterscheiden:

- *Einheit Y empfiehlt nicht negativ:* In diesem Fall revidiert die Einheit X ihren Glauben über die Einheit Y mit Hilfe des Revisionsfaktors r_K. Einheit Y wird also von Einheit X abgewertet.

- *Einheit Y empfiehlt negativ, nachdem Einheit Z negativ empfohlen hat:* Zuerst erhält die Einheit X die negative Empfehlung der Einheit Z und wertet wie im ersten Fall Einheit Y mit Hilfe des Revisionsfaktors r_K ab. Die nachfolgende negative Empfehlung der Einheit Y ändert nichts an dieser Abwertung. Sie bewirkt lediglich, dass auch Einheit Z von Einheit X abgewertet wird. Aus der Sicht der Einheit Y stimmt dieser Fall also mit dem ersten Fall überein.

- *Einheit Y empfiehlt negativ, bevor Einheit Z negativ empfohlen hat:* Dadurch, dass Einheit Y Einheit Z negativ empfiehlt, revidiert Einheit X ihren Glauben über Einheit Z. Nach dieser Glaubensrevision erachtet Einheit X Einheit Z als weniger normativ als zuvor. Erhält Einheit X in der Folge die negative Empfehlung der Einheit Z, so kommt daher bei der Abwertung der Einheit Y nicht mehr der Revisionsfaktor r_K der ersten beiden Fälle sondern der Faktor r'_K zum Einsatz. Die Abwertung der Einheit Z hat nämlich zu einer Verminderung ihrer Glaubwürdigkeit geführt. Für diesen Revisionsfaktor gilt $r'_K < r_K$. Somit fällt die Abwertung der Einheit Y schwächer als in den beiden ersten Fällen aus.

Die Betrachtung dieser drei Fälle zeigt, dass für Einheit Y das Ausstellen einer negativen Empfehlung sehr wohl eine Auswirkung auf sie selbst haben kann. Dies trifft genau dann zu, wenn sie vor der Einheit Z negativ empfiehlt. In diesem Fall führt die Beteiligung am Konflikt für die Einheit Y zu einer schwächeren Abwertung bei der Einheit X. Es gibt also einen Vorteil für diejenige Einheit, die zuerst empfiehlt. Diesen Vorteil nennen wir im Folgenden den *komparativen Vorteil des Erstempfehlens*.

Abbildung 7.8 zeigt eine Quantifizierung dieses komparativen Vorteils für eine beispielhafte Ausgangssituation[6]. Auf die x-Achse ist der priore Glaube der Einheit X über die Einheit Y aufgezeichnet. In y-Richtung trägt die Abbildung die Höhe der Abwertung der Einheit Y auf, die sich durch die Berücksichtigung der negativen Empfehlung der Einheit Z ergibt. Diese Höhe ergibt sich aus der Differenz des prioren und posterioren Typglaubens über die Einheit Y. Die durchgezogene Linie zeigt, wie stark die Einheit Y abgewertet wird, wenn sie nicht oder erst als Zweite negativ empfiehlt. Die gestrichelte Linie gibt die Abwertung für den Fall an, dass Einheit Y als Erste negativ empfiehlt. Die Differenz dieser beiden Kurven gibt den komparativen Vorteil des Erstempfehlens an. Er ist besonders hoch, wenn sich Einheit X vor Auftreten des Konflikts über den Typ der Einheit Y am unsichersten ist. Dies ist für Werte des prioren Typglaubens um 50% der Fall. Die Abbildung zeigt, dass es einen komparativen Vorteil des Erstempfehlens gibt, solange sich Einheit X über den Typ der Einheit Y unsicher ist. Diese Aussage trifft nicht nur

[6]Die Darstellung geht davon aus, dass Einheit X einen prioren Typglauben über die Normativität der Einheit Z von $p_X(N_Z) = 50\%$ besitzt. Außerdem schätzt Einheit X für den Transaktionskontext γ, bei dem es zum Konflikt zwischen den Einheiten Y und Z gekommen ist, die Wahrscheinlichkeit strategischer Einhaltung auf $p_n(\gamma) = 30\%$ und die des unbeabsichtigten Fehlverhaltens auf $p_u(\gamma) = 5\%$ ein.

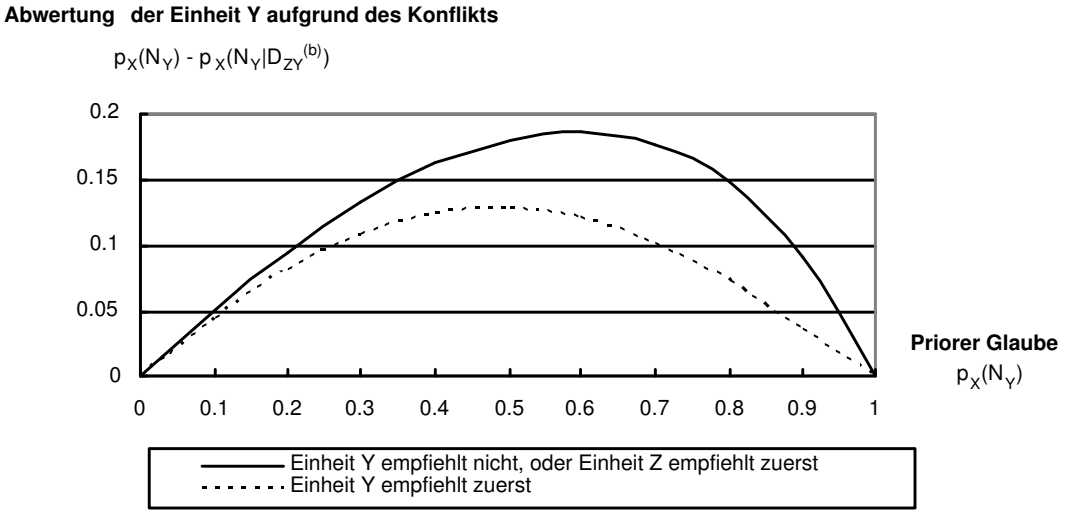

Abbildung 7.8: Der komparative Vorteil des Erstempfehlens

in der gewählten Ausgangssituation sondern allgemein zu. Unterschiedlich ist lediglich die Höhe des komparativen Vorteils. Er kann je nach Ausgangssituation höher oder niedriger ausfallen.

Wie beeinflusst das Erzielen des komparativen Vorteils den Nutzen der Einheit Y? Zunächst einmal bewirkt das Erstempfehlen lediglich, dass Einheit X Einheit Y nicht so stark abwertet. Damit führt der komparative Vorteil dazu, dass Einheit X eher zu Transaktionen mit der Einheit Y bereit ist. Durch das Erstempfehlen erhöht Einheit Y also ihre Chancen, in Zukunft Gelegenheiten zu Transaktionen mit der Einheit X zu erhalten. Den Nutzen, den sie erwartet, aus diesen Transaktionen zu ziehen, bezeichnen wir mit u_e. Er ist der Nutzen, den das Erzielen des komparativen Vorteils mit sich bringt.

Auf der anderen Seite verursacht das Ausstellen einer negativen Empfehlung die Kosten c_e. Sie entstehen daraus, dass eine negative Empfehlung ein Beweismittel ist und ihre Ausstellung somit das Ausführen einer kryptographischen Operation erfordert. Dies führt zu einer Beanspruchung der Batterie des Gerätes, auf dem sich die Einheit befindet, und letztendlich zur Notwendigkeit für den menschlichen Prinzipal, sein Gerät früher aufladen zu müssen. Wie stark wird jedoch die Batterie durch das negative Empfehlen beansprucht? Die folgenden Punkte zeigen, dass die Beanspruchung sehr gering ist:

- Die Messungen in [Kir05] zeigen, dass beim Ausstellen einer negativen Empfehlung auf realen Informationsgeräten lediglich 1 mJ verbraucht werden würden[7]. Heutzutage erhältliche Informationsgeräte verfügen über Batterien, deren Kapazität bei Laptops um die $2,2$ Ah (8 kJ) und bei PDAs um die $0,95$ Ah ($3,4$ kJ) liegt. Damit erhalten wir, dass über drei Millionen negative Empfehlungen ausgestellt werden können bevor ein Aufladen des Informationsgerät fällig wird. Anders herum gesagt, trägt das Ausstellen einer negativen Empfehlung nur unwesentlich zur Entladung der Batterie bei. Dieser Sachverhalt trifft insbesondere auch auf die zukünftigen Generationen von Informationsgeräten zu, bei denen aufgrund der zu erwartenden höheren Energieeffizienz der relative Batterieverbrauch für kryptographische Operationen noch geringer ausfallen wird.

[7]Diesen Wert erhält die Messung für den Fall, dass die zu signierende Aussage im Bereich von 100 Byte liegt. Dies entspricht der maximalen Größe der Aussage einer negativen Empfehlung.

- Der Inhalt einer negativer Empfehlung hängt nicht von ihrem Adressaten ab. Die Einheit Y muss damit für das Ausstellen der negativen Empfehlung unter Umständen keine kryptographische Operation ausführen. Dies ist dann der Fall, wenn sie zu einem früheren Zeitpunkt bereits die Einheit Z einer anderen Einheit W gegenüber negativ empfohlen hat. Dann reicht es für Einheit Y aus, die dabei ausgestellte negative Empfehlung bei sich lokal speichern und auf Anfrage der Einheit X zur Verfügung zu stellen. Wir erhalten damit, dass die Beteiligung an einem Konflikt und nicht das Ausstellen einer negativen Empfehlung selbst einen Energieverbrauch von $1mJ$ mit sich bringt. Dies führt zu einer weiteren Verminderung des relativen Batterieverbrauchs.

- Der Energieverbrauch für die Weitergabe der negativen Empfehlung liegen gemäß der Messungen aus [PS04] bei $0,2\ mJ$.[8] Somit ist der Aufwand für die Übertragung einer negativen Empfehlung ebenfalls sehr gering.

Diese Punkte zeigen, dass das Ausstellen einer negativen Empfehlung nur sehr unwesentlich zur Entladung der Batterie beiträgt. Daher sind die Kosten für das negative Empfehlen c_e um ein Vielfaches geringer als der Nutzen, den die Teilnahme an Transaktionen und der damit verbundene Austausch von benötigten Informationen mit sich bringt. Da der komparative Vorteil des Erstempfehlens gleichbedeutend mit mehr Gelegenheiten zu Transaktionen ist, ist der Nutzen u_e aus dem kooperativen Vorteil somit weitaus größer als die Kosten c_e für das negative Empfehlen.

Zusammenfassend haben wir gezeigt, dass das negative Empfehlen für die Einheit Y von Vorteil ist, wenn dies vor der negativen Empfehlung der Einheit Z geschieht und Einheit Y somit den komparativen Vorteil des Erstempfehlens erhält. Bei der bisherigen Analyse sind wir allerdings davon ausgegangen, dass Einheit Z unabhängig von den Überlegungen der Einheit Y negativ empfehlen wird. Wenn Einheit Z dies nicht tut, besitzt Einheit Y keinen Anreiz zum negativen Empfehlen. Es stellt sich damit die Frage, wie Einheit Y das Empfehlungsverhalten der Einheit Z antizipieren kann. An diese Fragestellung gehen wir im nachfolgenden Paragraphen mit Hilfe einer spieltheoretischen Betrachtung.

Das Spiel negativer Empfehlungen. Tabelle 7.2 zeigt eine Modellierung der Ausgangssituation als Spiel. Nach ihrem Konflikt haben sowohl Einheit Y als auch Einheit Z die Möglichkeit eine negative Empfehlung an die Einheit X auszustellen (E) oder darauf zu verzichten ($\neg E$). Entscheidet sich keine der beiden Einheiten für das negative Empfehlen, so erfährt Einheit X nicht von ihrem Konflikt. Infolgedessen wertet sie keine der beiden am Konflikt beteiligten Einheiten ab. Dies wird in der Modellierung dadurch berücksichtigt, dass für den Ausgang $(\neg E, \neg E)$ weder Einheit Y noch Einheit Z Einbußen ihres Nutzenniveaus hinnehmen müssen und ihre Auszahlungen somit null betragen. Entscheidet sich eine Einheit zum negativen Empfehlen, so fallen für sie die Kosten c_e des negativen Empfehlens an. Zugleich entstehen für den negativ Empfohlenen die Kosten c_k. Sie ergeben sich daraus, dass er durch die Abwertung in Zukunft von der Einheit X seltener als Transaktionspartner akzeptiert wird. Kommt es zum beidseitigen negativen Empfehlen, so erhält entweder Einheit Y oder Einheit Z den Vorteil des Erstempfehlens. Beide Einheiten haben aber keinen direkten Einfluss darauf, ob sie als Erste von der Einheit X aufgefordert werden, eine negative Empfehlung auszustellen. Daraus ergibt sich, dass bei einem Ausgang (E, E) beide Einheiten mit der gleichen Wahrscheinlichkeit den komparativen Vorteil, im Mittel also den halben komparativen Vorteil erhalten. Dies kommt in den Auszahlungen dadurch zum Ausdruck, dass der Term $\frac{u_e}{2}$ zum Nutzenniveau beider Einheiten addiert wird.

[8] Dieser Wert ergibt sich aus der Übertragung einer Nachricht mit 200 Byte Länge. Dies entspricht der Größe

Tabelle 7.2: Das Spiel negativer Empfehlungen

		Einheit Z	
		¬E	E
Einheit Y	¬E	$(0, 0)$	$(-c_k, -c_e)$
	E	$(-c_e, -c_k)$	$(-c_e - c_k + \frac{u_e}{2}, -c_e - c_k + \frac{u_e}{2})$

Die Tabelle zeigt das Spiel negativer Empfehlungen in seiner Normalform. Ein entscheidendes Charakteristikum des Spiels liegt darin, dass keiner der beiden Spieler bei der Wahl seiner Strategie weiß, ob sein jeweiliger Gegenüber bereits negativ empfohlen hat. Dies folgt direkt aus der Eigenschaft des Systemmodells, dass weder Einheit Y noch Einheit Z ihre Kommunikation mit der Einheit X gegenseitig einsehen können. Auch aus dem Einholen von Empfehlungen durch die Einheit X lässt sich nicht ableiten, ob der Gegenüber bereits negativ empfohlen hat oder nicht. Bei dem Spiel handelt es sich also um ein simultanes Spiel.

Analyse des Spiels negativer Empfehlungen. Bei dem Spiel negativer Empfehlungen handelt es sich um ein Koordinationsspiel: Ist Einheit Y bekannt, dass Einheit Z nicht negativ empfehlen wird, so ist es auch für sie von Vorteil, keine negative Empfehlung auszustellen. Dadurch vermeidet sie die Kosten c_e. Ist umgekehrt bekannt, dass Einheit Z eine negative Empfehlung ausstellt, wird auch die Einheit Y negativ empfehlen. Durch diese negative Empfehlung erhöht sie ihre Auszahlung um $\frac{u_e}{2} - c_e$. Das Spiel hat somit zwei Gleichgewichte in reinen Strategien, nämlich $(\neg E, \neg E)$ und (E, E).

Gemäß Abschnitt 4.1.1 ist für Koordinationsspiele charakteristisch, dass sie zudem ein labiles Gleichgewicht in gemischten Strategien besitzen. Dieses leiten wir für das Spiel negativer Empfehlungen wie folgt ab: Sei p_{eq} die Wahrscheinlichkeit dafür, dass Einheit Z negativ empfiehlt. Damit in dieser Wahrscheinlichkeit die Gleichgewichtsbedingung erfüllt ist, muss Einheit Y indifferent über ihre Handlungsalternativen sein. Wir erhalten also:

$$\begin{aligned} -c_k \cdot p_{eq} &= -c_e \cdot (1 - p_{eq}) + [-c_e - c_k + \frac{u_e}{2}] \cdot p_{eq} \\ 0 &= -c_e \cdot (1 - p_{eq}) + [-c_e + \frac{u_e}{2}] \cdot p_{eq} \\ 0 &= -c_e + \frac{u_e}{2} \cdot p_{eq} \\ p_{eq} &= 2 \cdot \frac{c_e}{u_e} \end{aligned} \tag{7.6}$$

Wenn beide Einheiten mit einer Wahrscheinlichkeit von p_{eq} negativ empfehlen, befinden sich ihre Strategien also im Gleichgewicht. Wegen $u_e \gg c_e$ ist diese Wahrscheinlichkeit sehr gering. Das Gleichgewicht ist insofern labil, als eine Abweichung von dieser Wahrscheinlichkeit zum Gleichgewicht in den reinen Strategien führt. Schätzt Einheit Y zum Beispiel die Wahrscheinlichkeit $p_Y(E_Z)$, dass Einheit Z negativ empfiehlt, höher als p_{eq} ein, so ist es für sie von Vorteil, die negative Empfehlung auszustellen. Zu welchem Gleichgewicht es letztendlich kommt, entscheidet somit die Einschätzung von $p_Y(E_Z)$:

- $p_Y(E_Z) = p_{eq}$: Einheit Y hält sich an die gemischte Gleichgewichtsstrategie.
- $p_Y(E_Z) > p_{eq}$: Einheit Y empfiehlt negativ. Dies führt zu einem Gleichgewicht in (E, E).

einer negativen Empfehlung, wenn die Steuerinformationen des Transportsystems mitgerechnet werden.

- $p_Y(E_Z) < p_{eq}$: Einheit Y empfiehlt nicht negativ. Wir erhalten ein Gleichgewicht in $(\neg E, \neg E)$.

In der bisherigen Analyse sind wir davon ausgegangen, dass sowohl Einheit Y als auch Einheit Z rationale Spieler sind. Die Situation ändert sich jedoch dadurch, dass normative Einheiten zum negativen Empfehlen vor-festgelegt sind. Dadurch gibt sich aus der Sicht der Einheit Y folgendes Bild: Sie weiß, dass die Normativität der Einheit Z eine hinreichende Bedingung dafür ist, dass sie negativ empfehlen wird. Ihr Typglaube $p_Y(N_Z)$ über Einheit Z gibt also eine untere Schranke für die Wahrscheinlichkeit $p_Y(E_Z)$ an. Aus der Analyse der Gleichgewichtsbedingungen erhalten wir $p_Y(N_Z) > p_{eq}$ als eine hinreichende Bedingung dafür, dass das negative Empfehlen für die Einheit Y von Vorteil ist. Da p_{eq} sehr klein ist, wird Einheit Y also immer negativ empfehlen, es sei denn, sie ist sich sehr sicher, dass die Einheit Z strategisch ist.

Wir fassen das Ergebnis der Analyse wie folgt zusammen: Eine strategische Einheit stellt im Falle eines Konflikts eine negative Empfehlung über ihren Transaktionspartner aus, solange es ihr nicht als unwahrscheinlich erscheint, dass er normativ ist. Wie verhält sich jedoch eine strategische Einheit, wenn sie sich ziemlich sicher ist, dass ihr Transaktionspartner strategisch ist? Hierüber macht die bisherige Analyse keine Aussage. Mit dieser Fragestellung beschäftigen wir uns daher im nachfolgenden Paragraphen.

Erweiterte Analyse basierend auf Meta-Glauben. Im Gegensatz zu normativen Einheiten sind strategische Einheiten nicht zum negativen Empfehlen vor-festgelegt. Sie stellen nur dann eine negative Empfehlung aus, wenn sie vermuten, dass die andere Einheit, mit der sie im Konflikt stehen, auch negativ empfehlen wird. Erscheint die Einheit Z als hinreichend normativ, so entscheidet sich eine strategische Einheit Y gemäß der vorigen Analyse zum negativen Empfehlen. Ist Einheit Z aber mit hoher Wahrscheinlichkeit strategisch, so muss Einheit Y antizipieren, wie sich Einheit Z entscheiden wird. Analog zur Einheit Y macht auch die Einheit Z, so sie strategisch ist, ihre Entscheidung abhängig von ihrem Typglauben $p_Z(N_Y)$ über die Einheit Y. Wenn die Einheit Y der Einheit Z als hinreichend normativ erscheint, so muss sie damit rechnen, von der Einheit Z negativ empfohlen zu werden. Wir erhalten somit eine Situation, in der die Einheit Y den Typglauben einschätzen muss, den Einheit Z über sie selbst hat. Da diese Einschätzung einen Glauben über den Glauben Anderer darstellt, nennen wir sie Meta-Glauben. Im Folgenden erweitern wir die Analyse durch die Berücksichtigung dieses Meta-Glaubens.

Aus der Sicht der Einheit Y ist ihre Einschätzung der Wahrscheinlichkeit $p_Y(E_Z)$ entscheidend, mit der Einheit Z sie bei Einheit X negativ empfiehlt. Einheit Y wird dann negativ empfehlen, wenn $p_Y(E_Z) > p_{eq}$ gilt. Die folgende Formel zeigt, wie diese Wahrscheinlichkeit $p_Y(E_Z)$ einzuschätzen ist:

$$\begin{aligned} p_Y(E_Z) &= p_Y(N_Z) \cdot p_Y(E_Z|N_Z) + p_Y(S_Z) \cdot p_Y(E_Z|S_Z) \\ &= p_Y(N_Z) + p_Y(S_Z) \cdot p_Y(E_Z|S_Z) \\ p_Y(\neg E_Z) &= p_Y(S_Z) \cdot p_Y(\neg E_Z|S_Z) \end{aligned} \tag{7.7}$$

Die Vereinfachung für $p_Y(E_Z)$ ergibt sich daraus, dass normative Einheiten zum negativen Empfehlen vor-festgelegt sind, also $p_Y(E_Z|N_Z) = 1$ gilt. Die Formeln zeigen, dass die Wahrscheinlichkeit $p_Y(\neg E_Z)$ für das Ausbleiben der negativen Empfehlung der Einheit Z einfacher einschätzen lässt. Der Übersichtlichkeit wegen wenden wir uns daher im Folgenden der Frage zu, wann eine strategische Einheit Y *nicht* negativ empfiehlt. Gemäß der vorigen Analyse muss dafür als Kriterium gelten:

$$p_Y(\neg E_Z) > 1 - p_{eq} \tag{7.8}$$

Im Folgenden leiten wir schrittweise notwendige Kriterien dafür ab, dass diese Ungleichung erfüllt ist:

- *Typglaube erster Stufe:* Aus der Formel 7.7 ergibt sich die Abschätzung $p_Y(\neg E_Z) < p_Y(S_Z)$. Sie basiert auf den Typglauben der Einheit Y. Aus der Abschätzung und dem Kriterium 7.8 lässt sich ableiten, dass $p_Y(S_Z) > 1 - p_{eq}$ (oder umgeformt $p_Y(N_Z) < p_{eq}$) gelten muss. Damit erhalten wir als notwendiges Kriterium, dass Einheit Y Einheit Z als ziemlich sicher strategisch ansieht. Eben dieses Ergebnis haben wir bereits bei der spieltheoretischen Analyse herausgefunden.

- *Typglaube zweiter Stufe:* Die Formel für $p_Y(\neg E_Z)$ lässt sich in eine andere Richtung abschätzen. Es gilt nämlich auch $p_Y(\neg E_Z) < p_Y(\neg E_Z|S_Z)$. Um eine Einschätzung für $p_Y(\neg E_Z|S_Z)$ zu finden, müssen wir lediglich die bisherigen Überlegungen auf Einheit Z anwenden. Wenn sie strategisch ist, orientiert sie genauso wie Einheit Y ihr Empfehlungsverhalten an ihrem Typglauben. Dabei verwendet sie das Kriterium 7.8 aus ihrer Sicht. Konkret bedeutet dies:

$$\begin{aligned} p_Y(\neg E_Z|S_Z) &= p_Y[1 - p_{eq} < p_Z(\neg E_Y)] \\ &= p_Y[1 - p_{eq} < p_Z(S_Y) \cdot p_Z(\neg E_Y|S_Y)] \end{aligned} \tag{7.9}$$

  Diese Einschätzung von $p_Y(\neg E_Z|S_Z)$ basiert auf Typglauben zweiter Stufe. Dies lässt sich in der Formel dadurch erkennen, dass Einheit Y den Typglauben der Einheit Z über sie selbst einschätzen muss. Rein formal wird dies aus dem Aufbau $p_Y[...p_Z...]$ ersichtlich. Die Formel erlaubt das Aufstellen einer zweiten notwendigen Bedingung dafür, dass Einheit Y nicht negativ empfiehlt. Sie wird analog zum Typglauben der ersten Stufe aus $p_Z(\neg E_Y) < p_Z(S_Y)$ abgeleitet. Daraus erhalten wir als notwendiges Kriterium für das Nichtempfehlen der Einheit Y:

$$\begin{aligned} 1 - p_{eq} &< p_Y(\neg E_Z|S_Z) < p_Y[1 - p_{eq} < p_Z(S_Y)] \\ \Longrightarrow \quad p_{eq} &> p_Y[1 - p_{eq} > p_Z(S_Y)] = p_Y[p_{eq} < p_Z(N_Y)] \end{aligned} \tag{7.10}$$

  Somit lautet das Kriterium, dass Einheit Y ziemlich sicher ist, dass Einheit Z sie als sehr strategisch ansieht.

- *Typglaube höherer Stufen:* Diese Betrachtung lässt sich bis zu einer beliebigen Höhe des Metaglaubens fortsetzen. Jeweils erhalten wir als notwendiges Kriterium für das Nichtempfehlen, dass Einheit Y und Z sich gegenseitig als strategisch ansehen und sich bis zu einem gewissen Rekursionsgrad jeweils sicher sind, dass der Andere dies auch so sieht.

Der Übersichtlichkeit wegen sind diese Schritte auf das Nichtempfehlen ausgerichtet gewesen. Hierfür wurden notwendige Kriterien hergeleitet. Wenn wir uns umgekehrt fragen, wann Einheit Y empfiehlt, werden aus diesen Kriterien hinreichende Bedingungen für das Ausstellen der negativen Empfehlung. Als Resultat der Analyse erhalten wir also, dass eine strategische Einheit Y *empfiehlt*, wenn eine der folgenden Bedingungen erfüllt ist:

- Einheit Y kann nicht ausschließen, dass Einheit Z normativ ist.

- Einheit Y kann nicht ausschließen, dass Einheit Z nicht ausschließen kann, dass Einheit Y normativ ist.

- Diese Art der Verschachtelung bildet in jeder beliebigen Tiefe eine hinreichende Bedingung für das negative Empfehlen.

Diese Bedingungen lassen sich wie folgt zusammenfassen: Einheit Y empfiehlt *immer* negativ. Die einzige Ausnahme davon besteht darin, dass die Einheiten Y und Z gemeinsames Wissen darüber haben, dass sie beide strategisch sind. Dieses gemeinsame Wissen lässt sich jedoch im Informationssystem nicht erreichen[9]. Damit erhalten wir, dass strategische Einheiten immer negativ empfehlen und dass dies unabhängig davon gilt, welche Einheit ihnen gegenübersteht.

Die Berücksichtigung von Metaglauben bei der spieltheoretischen Analyse entspricht den Untersuchungen des Ladenkettenspiels, die in Abschnitt 4.1.2 besprochen worden sind. Dort reicht es für den Markteintritt des potentiellen Konkurrenten nicht aus, dass er über die Schwäche des Monopolisten informiert ist. Darüber hinaus muss der Monopolist Kenntnis von diesem Wissen des potentiellen Konkurrenten haben, da er ansonsten versucht sein könnte, sich durch den Preiskampf als starker Monopolist darzustellen. Auch hier lässt sich diese Überlegung rekursiv fortführen. Das Ergebnis unserer Analyse des Spiels negativer Empfehlungen steht somit in direktem Zusammenhang mit den Untersuchungen zum Ladenkettenspiel.

Fazit. Die Vorschriften der Glaubensrevision sorgen durch den komparativen Vorteil des Erstempfehlers dafür, dass es einen indirekten Anreiz zum Ausstellen negativer Empfehlungen gibt. Dieser führt dazu, dass auch strategische Einheiten negativ empfehlen. Die spieltheoretische Analyse hat nicht nur gezeigt, dass sie dies tun, wenn sie ihren jeweiligen Gegenüber als möglicherweise normativ ansehen. Die erweiterte Analyse basierend auf Metaglauben hat zudem hervorgebracht, dass die Unsicherheit über den Glauben des jeweiligen Gegenübers ein ausreichender Grund für das negative Empfehlen ausstellt. Somit konnte gezeigt werden, dass sich auch strategische Einheiten aktiv am Empfehlungssystem beteiligen.

7.6.3 Bewertung

Der Einsatz transaktionaler Beweismittel zielt darauf, die soziale Kontrolle über das transaktionale Verhalten der Einheiten zu verstärken. Dies bedeutet konkret, dass Betrugsverhalten im Zuge einer Transaktion nicht nur vom jeweiligen Transaktionspartner sondern auch von anderen Einheiten als solches wahrgenommen wird. Dies wird durch die zwei folgenden Punkte gewährleistet:

- *Kopplung von Betrugsverhalten und Konflikten:* Der Systementwurf ist der Einschränkung des Systemmodells unterworfen, dass transaktionales Verhalten nur vom jeweiligen Transaktionspartner beobachtet werden kann. Kommt es zu Betrugsverhalten, so führt dies zu einem Konflikt zwischen den Transaktionspartnern. Andere Einheiten, die von diesem Konflikt wissen, können zwar nicht erkennen, welche der beiden Transaktionspartner tatsächlich betrogen hat. Sie sind aber in der Lage, aus der Häufung der Konfliktbeteiligung einer Einheit zu schließen, dass diese Einheit wohl der Urheber der Konflikte ist und daher wahrscheinlich strategisch ist. Dafür sorgt die Glaubensrevision, die jede Einheit für sich lokal durchführt. Die Grundlage dieser Überlegung ist, dass strategische Einheiten im Mittel an mehr Konflikten beteiligt sind als normative Einheiten. Dies ergibt sich direkt daraus,

[9] Dies gilt auch dann, wenn aufgrund von Typbeweisen beide Einheiten als strategisch erkannt worden sind. Dann sind sich nämlich Einheit Y und Z nicht sicher, dass ihr jeweiliger Gegenüber von den jeweiligen Typbeweisen erfahren hat. Selbst wenn sie sich dies gegenseitig bestätigen würden, hätten sie dennoch kein gemeinsames Wissen von ihrem Typ. Dies zeigt das Problem des koordinierten Angriffs aus Abschnitt 4.2.

dass nur strategische Einheiten die Absicht zu Betrugsverhalten und damit die Absicht zu Konflikten haben können.

- *Kopplung von Konflikten und negativen Empfehlungen:* Die bisherige Überlegung beruht darauf, dass es einen verlässlichen Mechanismus gibt, durch den die Einheiten von Konflikten erfahren. Diese Verlässlichkeit ist nur dann gegeben, wenn eine Kopplung zwischen dem Auftreten von Konflikten und die Benachrichtigung darüber in Form einer negativen Empfehlung besteht.

Die Analyse der vorangegangenen beiden Abschnitte beschäftigte sich mit dem zweiten Punkt, der Kopplung von Konflikten und negativen Empfehlungen. Die Ergebnisse der Analyse fassen wir im Folgenden in zwei Schritten zusammen. Sie zeigen, dass das Auftreten von Konflikten sowohl *notwendig* als auch *hinreichend* für das Ausstellen einer entsprechenden negativen Empfehlung ist.

In Abschnitt 2.4.2 haben wir gefordert, dass Empfehlungen verfügbar und wahrheitsgemäß sind. Diese beiden Forderungen sind nichts anderes als die beiden oben genannten Kriterien: Negative Empfehlungen sind insofern *wahrheitsgemäß*, als aus ihnen auf einen Konflikt gefolgert werden kann. Dies besagt das notwendige Kriterium. Die negativen Empfehlungen sind darüber hinaus *verfügbar*, solange das hinreichende Kriterium erfüllt ist. Es stellt sicher, dass jeder Konflikt zu entsprechenden negativen Empfehlungen führt.

Konflikte als notwendiges Kriterium für negative Empfehlungen. Hierbei ist zu zeigen, dass aus dem Erhalt einer negativen Empfehlung auf das Auftreten eines Konflikts in der entsprechenden Transaktion geschlossen werden kann.

Das Empfehlungssystem fordert, dass ein entsprechender Vertrag zwingendermaßen die negative Empfehlung begleitet. Aus diesem Vertrag lässt sich erkennen, dass in der Tat eine Transaktion zwischen den beiden Einheiten stattgefunden hat. Angenommen, in dieser Transaktion wäre es nicht zum Konflikt gekommen, dann wäre der negativ Empfohlene im Besitz einer Quittung, die die negative Empfehlung widerlegt. Der Aussteller der negativen Empfehlung müsste also davon ausgehen, dass seine inkonsistenten Festlegungen von den anderen Einheiten wahrgenommen werden würden. Hierfür sorgen eine entsprechende Selbstempfehlung und nachfolgende Typbeweise. Um zu vermeiden, als strategisch erkannt zu werden, würde also eine Einheit auf das Ausstellen einer negativen Empfehlung unter diesen Umständen verzichten. Damit ist die obige Annahme widerlegt und wir erhalten, dass es tatsächlich in der Transaktion zum Konflikt gekommen ist.

Konflikte als hinreichendes Kriterium für negative Empfehlungen. Wir müssen zeigen, dass bei Auftreten eines Konflikts beide Transaktionspartner *fähig* und *willens* sind, eine entsprechende negative Empfehlung auszustellen.

Eine Einheit ist zum negativen Empfehlen fähig, wenn sie einen entsprechenden Vertrag besitzt und zuvor keine Quittung dem Transaktionspartner übergeben hat. Beide Punkte werden durch die Analyse zur Ausstellung von Verträgen und Quittungen in Abschnitt 7.6.1 bestätigt. **(1)** Betrugsverhalten ist für die Transaktionspartner nur nach dem beidseitigem Ausstellen von Verträgen von Interesse. Dies ergibt sich daraus, dass nur dann die Vorteile aus der Aktionsausführung des Gegenübers gezogen werden können. **(2)** Eine Einheit stellt erst dann eine Quittung aus, wenn sowohl ihr Transaktionspartner als auch sie selbst ihre jeweiligen Aktionen ausgeführt haben. Die Analyse zeigt, dass nicht davon auszugehen ist, dass nach der Ausstellung dieser Quittung der

eigene Transaktionspartner betrügt. Er würde sich dadurch im Empfehlungssystem keinen Vorteil verschaffen und sich zudem die Möglichkeit zukünftiger Transaktionen mit der betrogenen Einheit verbauen.

Bei Auftreten eines Konflikts ist eine Einheit nicht nur fähig zum negativen Empfehlen. Darüber hinaus ist sie dazu willig, wie das Spiel negativer Empfehlungen und dessen Analyse in Abschnitt 7.6.2 zeigt. Dieser Sachverhalt ergibt sich aus dem komparativen Vorteil derjenigen Einheit, die als Erste negativ empfiehlt. Diesen Vorteil sehen sich auch strategische Einheiten genötigt auszunutzen, da sie darüber unsicher sind, ob ihr Transaktionspartner normativ ist und damit negativ empfehlen wird. Diese Unsicherheit wird dadurch gesteigert, dass der Transaktionspartner, selbst wenn er strategisch ist, unter Umständen einen selbst negativ empfehlen könnte. Als Ergebnis erhalten wir, dass nicht nur normative sondern auch strategische Einheiten willens zum negativen Empfehlen sind.

Fazit. Betrugsverhalten führt unweigerlich zu einem Konflikt mit dem eigenen Transaktionspartner. Durch den Einsatz transaktionaler Beweismittel erreicht der Systementwurf, dass dieser Konflikt auch von unbeteiligten Einheiten wahrgenommen wird. Dadurch, dass strategische Einheiten eher an Konflikten beteiligt sind als normative Einheiten, dient die Wahrnehmung von Konflikten als Grundlage dafür, den Typ anderer Einheiten einzuschätzen. Damit haben wir unser Ziel der verteilten Vertrauensbildung erreicht: Die soziale Kontrolle zwischen den Einheiten wird verschärft, weil die Einheiten gegenseitig von ihren Erfahrungen profitieren.

7.7 Zusammenfassung

In diesem Kapitel haben wir uns mit der Verschärfung der sozialen Kontrolle befasst, die durch den Austausch der Transaktionserfahrungen der einzelnen Einheiten ermöglicht wird. Die sich ergebende Vertrauensbildung ist insofern verteilt, als in sie nicht nur die eigenen Transaktionserfahrungen sondern auch die der anderen Einheiten einfließen. Die Grundlage hierfür bildet der Einsatz von *Beweismitteln*, auf deren Konzept wir zunächst eingegangen sind. Dadurch, dass Festlegungen in Beweismitteln auf nicht-abstreitbare Weise eingegangen werden, entstehen Möglichkeiten zur glaubwürdigen Signalisierung. In diesem Kapitel haben wir uns mit dem Einsatz von transaktionalen Beweismitteln beschäftigt, deren Aussagen sich auf das Transaktionsverhalten einer Einheit beziehen. Der Kreislauf der Vertrauensbildung wurde entsprechend erweitert, um die Ausstellung, Verteilung und Bewertung dieser transaktionalen Beweismittel zu ermöglichen.

Wir haben Verträge und Quittungen als zwei Arten von Beweismitteln identifiziert, deren Ausstellung im Rahmen einer Transaktion sinnvoll ist. Ihr Austausch führt zu einer Erweiterung des *Transaktionsprotokolls* zu einem Sechs-Wege Protokoll. Hinzu kommt als dritte Art transaktionaler Beweismittel die negative Empfehlung, die im Zuge des Empfehlungssystem eingeführt wurde. Der Entwurf des *Empfehlungssystems* wurde von der Frage geleitet, welche Möglichkeiten der glaubwürdigen Signalisierung der Einsatz transaktionaler Beweismittel eröffnet. Die drei identifizierten Empfehlungsarten stellen allesamt eine Form der indirekten Signalisierung dar. Verträge, Quittungen und negative Empfehlungen bilden jeweils als begleitende transaktionale Beweismittel die Grundlage für negativen Empfehlungen, Selbstempfehlungen und Typbeweise. Eine Sonderrolle nimmt die negative Empfehlung ein, da sich ihr Aussteller in einem entsprechenden Beweismittel auf ihre Aussage festlegt. Im Zuge des Empfehlens kommt es aufgrund der indirekten Signalisierung zu einer Verteilung und Weitergabe der Beweismittel zwischen den Einheiten. Entsprechende Vorschriften für das Einholen und Zusammenstellen von Empfehlungen

sind vorgestellt worden.

Die Berücksichtigung von Empfehlungen führte zu einer Erweiterung der *Glaubensbildung*. Negative Empfehlungen zeigen einen Konflikt zwischen zwei Einheiten an. Das inkonsistente Verhalten der Einheiten, die ohne Auftreten eines Konflikts dennoch negativ empfehlen, wird mit Hilfe von Selbstempfehlungen und Typbeweisen erkannt. Diese führen nicht nur dazu, dass die inkonsistent handelnde Einheit als strategisch angesehen wird. Darüber hinaus wird diejenige Einheit rehabilitiert, die die inkonsistente negative Empfehlung widerlegt hat. Die entsprechenden Revisionsvorschriften für die Bewertung von Konflikten und Rehabilitierungen wurden auf probabilistisch fundierte Weise abgeleitet und anschaulich interpretiert.

Die Wissensverwaltung einer Einheit wurde zu einer kombinierten *Beweismittel- und Wissensverwaltung* ausgebaut, um den Anforderungen der verteilten Vertrauensbildung genüge zu tun. Die Beweismittel und das Wissen, über das eine Einheit aufgrund eingeholter Empfehlungen und ihrer Teilnahme an Transaktionen verfügt, sind nicht nur lokal abzulegen. Darüber hinaus muss die Ablageverwaltung auch einen assoziativen Zugriff darauf ermöglichen. Dabei darf die geforderte Funktionalität nicht zu einem unkontrollierten Speicherverbrauch führen. Wir haben anhand einer möglichen Implementierung der Beweismittel- und Wissensverwaltung dargelegt, wie diese Anforderungen umzusetzen sind.

Abschließend haben wir in einer *Analyse* gezeigt, dass sich strategische Einheiten aus ihrem eigenen Interesse an die Vorgaben des Systementwurfs halten. Dies betrifft sowohl das Ausstellen von Verträgen und Quittungen im Zuge von Transaktionen sondern auch das aktive Empfehlen, wie in der Analyse des Spiels negativer Empfehlungen deutlich geworden ist. Dadurch stellen strategische Einheiten nicht nur ihre Transaktionserfahrungen anderen Einheiten zur Verfügung. Darüber hinaus beteiligen sie sich an den sozialen Kosten, die durch den Einsatz transaktionaler Beweismittel entstehen. Die Selbstdurchsetzung des Systementwurfs ermöglicht, dass das Auftreten eines Konflikts zwischen zwei Transaktionspartnern sowohl eine notwendige als auch eine hinreichende Bedingung für das Ausstellen einer negativen Empfehlung ist. Da damit die Zahl der Konflikte, an denen eine Einheit beteiligt ist, Aufschluss über ihren Typ gibt, werden die Folgekosten für Betrugsverhalten nachhaltig erhöht. Durch den Einsatz transaktionaler Beweismittel erreichen wir also unser Ziel, die soziale Kontrolle zwischen den Einheiten zu verschärfen.

Kapitel 8

益者三友 损者三友
友直 友谅 友多闻 益矣
友便辟 友善柔 友便佞 损矣

"Es gibt je drei Arten von nützlichen und schädlichen Freunden: Freundschaft mit denen, die rechtschaffen, aufrichtig und gut informiert sind, ist nützlich. Hingegen ist Freundschaft mit denen, die unterwürfig sind, ihre Prinzipien der jeweiligen Situation anpassen und raffiniert reden, schädlich."
(Gespräche und Aussprüche des Konfuzius, 16.4)

Erweiterung um soziale Beweismittel

Durch den Einsatz transaktionaler Beweismittel wurde im vorangehenden Kapitel die soziale Kontrolle zwischen den Einheiten verschärft. Strategische Einheiten werden als solche erkannt, da ihre wiederholte Betrugsabsicht im Rahmen des Empfehlungssystems unweigerlich auch von unbeteiligten Einheiten wahrgenommen wird. Dies wird durch negative Empfehlungen erreicht, die eine beherrschende Rolle im Empfehlungssystem innehaben. Hingegen sind die Möglichkeiten positiver Empfehlungen beschränkt, da Selbstempfehlungen nur zur Widerlegung negativer Empfehlungen eingesetzt werden. Normative Einheiten benötigen jedoch eine Möglichkeit zu positiven Empfehlungen, um sich durch Selbstempfehlungen gezielt von strategischen Einheiten absetzen zu können. Der Entwurf muss daher um eine weitere Art von Beweismittel erweitert werden, mit denen sowohl aussagekräftige als auch glaubwürdige Selbstempfehlungen ausgestellt werden können.

Mit dieser Aufgabe beschäftigt sich dieses Kapitel. Zu diesem Zweck werden in Abschnitt 8.1 soziale Beweismittel als Grundlage von Selbstempfehlungen eingeführt und der Kreislauf der Vertrauensbildung entsprechend erweitert. Die Aspekte des Einsatzes sozialer Beweismittel werden in Abschnitt 8.2 erörtert. Die Schlüsselfrage besteht darin, wie soziale Beweismittel in die Glaubensbildung einzubeziehen sind. Damit beschäftigt sich Abschnitt 8.3. Dies ermöglicht in Abschnitt 8.4 das Aufstellen von Vorschriften, wann Einheiten Beziehungen eingehen und dies mit Beweismitteln dokumentieren. Zudem wird untersucht, wie sich strategische Einheiten im Bezug auf diese Vorschriften verhalten.

8.1 Einführung

Im Folgenden geben wir eine Übersicht der Funktionsweise der verteilten Vertrauensbildung unter Zuhilfenahme sozialer Beweismittel. Hierfür gehen wir zunächst in Abschnitt 8.1.1 auf das Kon-

zept der sozialen Bindungen ein, das den sozialen Beweismitteln zugrunde liegt. Anschließend stellen wir in Abschnitt 8.1.2 dar, wie der Kreislauf der Vertrauensbildung zu erweitern ist, wenn soziale Beweismittel zum Einsatz kommen.

8.1.1 Konzept der sozialen Bindungen

Durch die Möglichkeit zur Kooperation entwickelt sich im Informationssystem ein soziales Gefüge. Dies kommt implizit darin zum Ausdruck, dass sich jede Einheit ihren eigenen Glauben über den Typ Anderer bildet. In diesem Abschnitt gehen wir einen Schritt weiter, indem wir soziale Bindungen einführen, die das soziale Gefüge explizit machen. Beweismittel, die diese Bindungen festhalten, eignen sich hervorragend als Grundlage von Selbstempfehlungen, da der Empfehler dadurch seine Stellung im sozialen Gefüge Anderen vermitteln kann.

Das Konzept der sozialen Bindungen stellen wir wie folgt vor: Zunächst untersuchen wir den Entwurfsraum von sozialen Bindungen. Anschließend identifizieren wir Typ-Bürgschaftsbeziehungen, die durch Beweismittel belegt sind, als eine viel versprechende Art der sozialen Bindung. Im Entwurf der nachfolgenden Abschnitte werden diese Bürgschaften als soziale Beweismittel eingesetzt.

Entwurfsraum. Unter einer *sozialen Bindung* verstehen wir eine Beziehung zwischen zwei oder mehreren Einheiten, die anderen Einheiten gegenüber mitgeteilt werden kann. Soziale Bindungen zielen also darauf, das soziale Gefüge, das implizit durch den Typglauben der einzelnen Einheiten gegeben ist, explizit zu machen.

Diese Definition der sozialen Bindung lässt einige Freiheitsgrade offen. Wir erhalten also einen Entwurfsraum von sozialen Bindungen, wie er in [OFN04] dargestellt ist. Bei den Freiheitsgraden handelt es sich um die folgenden Dimensionen:

1. *Bindung zwischen wie vielen Einheiten?* Laut Definition sind an einer sozialen Bindung zwei oder mehrere Einheiten beteiligt. Ein Spezialfall stellt daher die *paarweise* Bindung dar, in der sich genau zwei Einheiten aneinander binden.

2. *Symmetrische oder asymmetrische Bindung?* Die Einheiten, die sich aneinander binden, müssen nicht notwendigerweise dieselben Pflichten und Rechte besitzen, wie dies bei einer *symmetrischen* Bindung der Fall ist. Im Gegensatz dazu ist auch eine *asymmetrische* Bindung denkbar, in der die Rollen der Einheiten unterschiedlich sind.

3. *Was bedeutet die Bindung?* Dies ist die zentrale Frage einer sozialen Bindung. Sie entscheidet darüber, wie die Kenntnis über die sozialen Bindungen Anderer in die eigene Glaubensbildung einfließen muss und unter welchen Umständen soziale Bindungen eingegangen werden.

4. *Wie wird eine Bindung eingegangen?* Wenn sich Einheiten einig sind, sich aneinander zu binden, bedürfen sie einen Mechanismus, mit dem sie diese Bindung eingehen können. Es gibt hierbei zwei prinzipielle Möglichkeiten: **(1)** Zum Eingehen einer sozialen Bindung reicht eine Absprache, die zwischen den Einheiten getroffen wird. **(2)** Die Einheiten belegen ihre soziale Bindung durch das Ausstellen entsprechender Beweismittel. Da Beweismittel laut Abschnitt 7.1.1 immer einseitige Festlegungen sind, muss also jede der Einheiten ein solches Beweismittel ausstellen. Wir nennen es ein *soziales Beweismittel*, da darin die soziale Bindung zum Ausdruck kommt.

5. *Wie erfahren Andere von der Bindung?* Ist eine Bindung durch entsprechende Beweismittel belegt, so genügt es, diese Beweismittel anderen Einheiten zukommen zu lassen. Schwieriger ist es, wenn eine Bindung durch eine einfache Absprache zustande gekommen ist. In diesem Fall muss bei allen Einheiten, die an dieser Bindung beteiligt sind, nachgefragt werden, um sich der Existenz der Bindung zu vergewissern. Dies ist nicht nur aufwändig, sondern in Umgebungen wie dem Ad-hoc-Netz des Campus-Szenarios, in dem überwiegend nur mit einer kleinen Zahl von Einheiten kommuniziert werden kann, nicht durchführbar.

6. *Wie wird eine Bindung aufgelöst?* Auch hier ist die Art der Bindung zu unterscheiden. Kommt sie durch einfache Absprache zustande, so genügt ein Widerruf einer der beteiligten Einheiten dazu, dass die Bindung aufgelöst ist. Dieser Widerruf muss den anderen Einheiten nicht mitgeteilt werden, sondern er kann auch lokal erfolgen. Dies liegt daran, dass andere Einheiten sowieso bei allen Beteiligten nachfragen müssen, um sich von der Existenz der Bindung zu überzeugen. Ist die Bindung hingegen durch Beweismittel belegt, so ergeben sich zwei Möglichkeiten. Die Einheit, die aus der Bindung austreten will, könnte dies zwar durch das Ausstellen eines widerrufenden Beweismittels kundtun. Allerdings kann sie dann nicht sichergehen, dass andere Einheiten von ihrem Widerruf erfahren. Es ist also denkbar, dass trotz Widerruf die Bindung nicht effektiv aufgelöst ist. Daher bietet sich als zweite Möglichkeit an, soziale Bindungen mit einem Gültigkeitszeitraum zu versehen. Innerhalb dieses Zeitraums kann die Bindung zwar nicht aufgelöst werden. Wird eine Bindung aber ungültig, so gilt sie ohne Zutun als aufgelöst. Wird diese zweite Möglichkeit angewandt, so ist bei der Definition des Gültigkeitszeitraums eine Abwägung zu treffen: Ist der Zeitraum zu kurz, müssen Bindungen durch das erneute Ausstellen von Beweismitteln immer wieder erneuert werden. Ist der Zeitraum hingegen zu lang, so laufen die beteiligten Einheiten die Gefahr, dass sie die Bindung zu einem späteren Zeitpunkt bereuen aber nicht auflösen können.

Die Untersuchung dieser Fragen zeigt, dass es unterschiedliche Arten von sozialen Bindungen geben kann. Im *Buddy-System* werden zum Beispiel Bindungen durch einfache Absprache getroffen[1]. Im nachfolgenden Paragraphen schlagen wir einen anderen Weg ein, indem wir eine Art der sozialen Bindung vorstellen, die auf sozialen Beweismitteln beruht.

Die Typ-Bürgschaft als soziales Beweismittel. Um eine angemessene Art der sozialen Bindung zu finden, müssen wir uns fragen, welches Ziel wir mit dem Einsatz sozialer Bindungen verfolgen. Soziale Bindungen sollen Selbstempfehlungen ermöglichen, die sowohl aussagekräftig als auch glaubwürdig sind, damit sich normative Einheiten von strategischen Einheiten abgrenzen können. Damit erhalten wir zwei Anforderungen an die sozialen Bindungen:

- *Aussagekraft:* Aus dem Eingehen einer sozialen Bindung muss eine Aussage über die Normativität der beteiligten Einheiten ableitbar sein. Nur dann ist es sinnvoll, soziale Bindungen als Grundlage von Selbstempfehlungen zu verwenden. Anders gesprochen soll die Beteiligung an einer sozialen Bindung ein *Signal* für die eigene Normativität darstellen.

- *Glaubwürdigkeit:* Die Aussagekraft dieses Signals ist an seine Glaubwürdigkeit gebunden. Das Signal ist nur dann glaubwürdig, wenn Einheiten unterschiedlichen Typs nicht den-

[1]Die Problematik, die sich hierbei ergibt, ist in der Besprechung des Entwurfsraums sozialer Bindungen nur kurz gestreift worden. Sie wird in dieser Arbeit nicht weiter verfolgt, da der Entwurfsteil auf sozialen Beweismitteln und nicht auf einfachen Absprachen beruht. Eine detaillierte Untersuchung der Problematik findet sich in [OFN04, FOKR04].

selben Zugang zu sozialen Bindungen besitzen. Die Kernidee hierbei besteht darin, dass aneinander gebundene Einheiten füreinander *bürgen*. Damit fällt das Verhalten einer Einheit auf die an ihr gebundenen Einheiten zurück. In diesem Zusammenhang sprechen wir vom Eingehen einer sozialen Bindung als *Investition*. Sie zahlt sich nur dann aus, wenn sich der Gebürgte kooperativ verhält. Dadurch wird der Zugang von strategischen Einheiten zu sozialen Bindungen eingeschränkt[2].

Basierend auf diesen Vorüberlegungen entwerfen wir die *Typ-Bürgschaftsbeziehung* (oder kurz Bürgschaftsbeziehung) als spezielle Art der sozialen Bindung wie folgt:

1. *Paarweise Bindung:* Die Beschränkung auf paarweise Bindungen stellt keine eigentliche Einschränkung dar. Dies ergibt sich daraus, dass aus einer Reihe von paarweisen Bindungen de facto mehrseitige Bindungen zustande kommen können. Auf der anderen Seite haben paarweise Bindungen den Vorteil, dass das Eingehen der Bindung lediglich eine bilaterale Koordination erfordert. Dies ist im Hinblick auf den Einsatz von Beweismitteln von Vorteil.

2. *Symmetrische Bindung:* Eine Bürgschaftsbeziehung besteht immer darin, dass ein Paar von Einheiten gegenseitig füreinander bürgt. Damit nehmen die beteiligten Einheiten symmetrische Rollen ein. Wir sprechen daher auch von den beiden *Partnern* der Bürgschaftsbeziehung. Eine asymmetrische Bindung wird nicht in Betracht gezogen, da sie strategischen Einheiten ermöglichen würde, einseitige Bürgschaften für normative Einheiten einzugehen. Damit würde strategischen Einheiten entgegen unseres Zieles der Zugang zu sozialen Bindungen erleichtert.

3. *Bedeutung:* In einer Bürgschaftsbeziehung bescheinigen sich die beiden Partner gegenseitig, dass sie beide denselben Typ besitzen. Normative Einheiten werden also nur mit denjenigen Einheiten eine Bürgschaft eingehen, die ihnen als normativ erscheinen. Diese Festlegung der Bedeutung von Bürgschaftsbeziehungen erfüllt die oben genannten Anforderungen. **(1)** Eine Einheit kann ihre Normativität signalisieren, indem sie eine Einheit als Bürgen besitzt, die als normativ erscheint. **(2)** Das Eingehen einer Bürgschaftsbeziehung stellt insofern eine Investition dar, als das Verhalten des Partners auf die Wahrnehmung seiner Normativität und damit indirekt auf die Wahrnehmung der eigenen Normativität abfärbt.

4. *Eingang durch Ausstellung von Beweismitteln:* Ein Paar von Einheiten stellt eine Bürgschaftsbeziehung her, indem sie sich gegenseitig ein entsprechendes soziales Beweismittel ausstellen. Die Aussage jedes dieser beiden Beweismittel besteht in dem Bekenntnis dazu, für den Typ des Partners zu bürgen. Wir sprechen daher in der Folge von solchen Beweismitteln kurz als Bürgschaft. Damit ist die Bürgschaft nach den drei Arten der transaktionalen Beweismittel die vierte Art von Beweismitteln, die im Entwurf zum Einsatz kommt.

5. *Verbreitung durch Selbstempfehlungen:* Eine Einheit ist in der Lage, sich selbst zu empfehlen, indem sie die Beweismittel, die sie von ihren Bürgen erhalten hat, an andere Einheiten weitergibt. Bürgschaften erlauben somit eine indirekte Signalisierung im Sinne des Abschnitts 7.1.1.

6. *Automatische Auflösung durch Angabe des Gültigkeitszeitraums:* Bürgschaftsbeziehungen beziehen sich immer nur auf einen bestimmten Zeitraum. Wird nach dem Ablauf der Bürgschaftsbeziehung von beiden Partnern eine Erneuerung der sozialen Bindung erwünscht, so

[2]Dass dies tatsächlich der Fall ist, werden wir in der abschließenden Betrachtung dieses Kapitels in Abschnitt 8.4.2 zeigen.

müssen sie erneut entsprechende Bürgschaften mit einem neuen Gültigkeitszeitraum ausstellen.

Auf dem ersten Blick erscheint es so, dass Bürgschaftsbeziehungen transitiv sind. Dass dem jedoch nicht so ist, zeigt die folgende Überlegung: Nehmen wir an, dass die Einheiten X und Y sowie die Einheiten Y und Z jeweils eine Bürgschaftsbeziehung haben. Besitzen damit die Einheiten X und Z implizit auch eine Bürgschaftsbeziehung? Eine Einheit X tritt in eine Bürgschaftsbeziehung, wenn sie von der Normativität ihres Partners Y hinreichend überzeugt ist. Formal halten wir dies durch B_X: N_Y fest, wobei das B auf den Glauben (engl.: belief) der Einheit X verweist[3]. Aufgrund der beiden Bürgschaftsbeziehungen erhalten wir also die vier Aussagen B_X: N_Y, B_Y: N_X, B_Y: N_Z und B_Z: N_Y. Wenn Bürgschaftsbeziehungen transitiv wären, so müsste aus diesen Aussagen auch B_X: N_Z und B_Z: N_X ableitbar sein. Dies ist aber nicht der Fall. Hierfür wäre nämlich die Hilfsaussage vonnöten, dass normative Einheiten sich nie täuschen, das heißt $(N_Y \wedge B_Y{:}N_Z) \rightarrow N_Z$. Diese trifft jedoch nicht zu. Wir erhalten also, dass Bürgschaftsbeziehungen *intransitiv* sind und in der Glaubensbildung entsprechend zu behandeln sind.

8.1.2 Soziale Beweismittel im Kreislauf der Vertrauensbildung

Der Einsatz von Bürgschaften als soziale Beweismittel führt dazu, dass der Kreislauf der Vertrauensbildung erweitert wird. Im Folgenden geben wir eine Übersicht über die invasive Anwendung sozialer Beweismittel. Dabei zeigen wir analog zu Abschnitt 7.1.2 zwei verschiedene Sichten auf den Kreislauf, nämlich eine funktionszentrische und eine datenzentrische Sicht. Abschließend zeigen wir anhand eines Beispiels den Lebenszyklus sozialer Beweismittel auf und verdeutlichen dadurch, welche Aspekte der Entwurf im Umgang mit sozialen Beweismitteln berücksichtigen muss.

Funktionszentrische Sicht. Bürgschaftsbeziehungen haben zwei Anknüpfungspunkte zum bisher vorgebrachten Kreislauf der verteilten Vertrauensbildung. Einerseits ist der Typglaube einer Einheit entscheidend dafür, für wen sie bereit ist zu bürgen. Andererseits führt die Bürgschaftsbeziehung zum Besitz eines entsprechenden sozialen Beweismittels, das in die Beweismittel- und Wissensverwaltung Eingang findet. Abbildung 8.1 gibt eine funktionszentrische Sicht auf eine Erweiterung des Kreislaufs, die diese beiden Punkte berücksichtigt. Dabei ist zu beachten, dass eine Einheit in ihrer Beweismittel-Basis nicht die Bürgschaft, die sie selbst ausgestellt hat, sondern die ihres Partners der Bürgschaftsbeziehung ablegt.

Aus der Abbildung geht unmittelbar hervor, dass für den Einsatz sozialer Beweismittel Anpassungen beziehungsweise Erweiterungen vonnöten sind. Sie müssen die folgenden Fragen beantworten:

- Wann geht eine Einheit Bürgschaftsbeziehungen ein?
- Wie läuft der Eingang in die Bürgschaftsbeziehung und der damit verbundene Austausch der Bürgschaften ab?
- Wie muss die Beweismittel- und Wissensverwaltung erweitert werden, um mit Bürgschaften umgehen zu können?

[3]Genauer lässt sich B_X: N_Y definieren als $p_X(N_Y) > p_\sigma$, wobei p_σ der Schwellwert des Typglaubens ist, bei dem von einem solchen hinreichenden Glauben gesprochen werden kann. Die Festlegung dieses p_σ im Rahmen des Entwurfs wird in Abschnitt 8.4.1 angegangen.

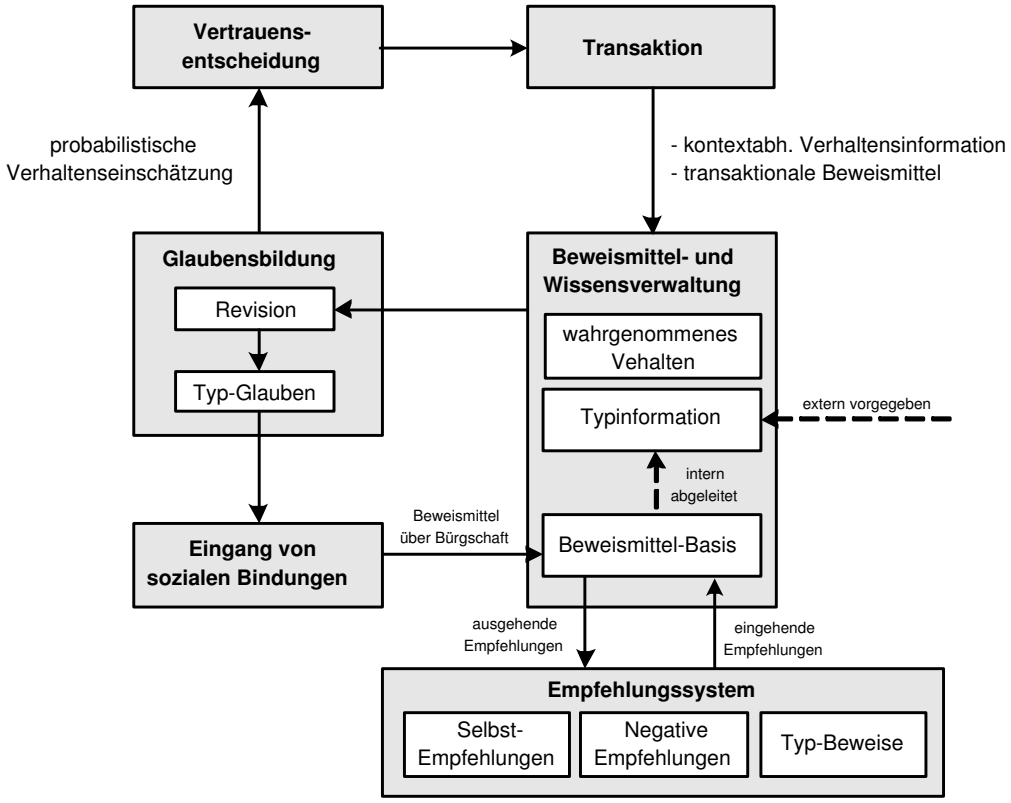

Abbildung 8.1: Erweiterung des Kreislaufs der verteilten Vertrauensbildung um soziale Beweismittel (funktionszentrische Sicht)

Des Weiteren ergeben sich einige weitere Betätigungsfelder, wenn wir uns vor Augen führen, dass die Bürgschaften in der Beweismittel- und Wissensverwaltung zur Verfügung stehen:

- Wie gehen Bürgschaften in das Empfehlungssystem ein?
- Wie ist in der Glaubensbildung das Wissen um die Bürgschaft zweier anderer Einheiten zu bewerten?

Auf diese insgesamt fünf Fragen werden in den Abschnitten 8.2, 8.3 und 8.4 Antworten gegeben.

Datenzentrische Sicht. Die Auswirkungen, die durch den Einsatz sozialer Beweismittel entstehen, werden besonders deutlich, wenn wir die datenzentrische Sicht des Kreislaufs der Vertrauensbildung einnehmen.

Abbildung 8.2 zeigt diese Sicht. Durch den Einsatz von Bürgschaftsbeziehungen wird eine zweite Möglichkeit geschaffen, dass sich der Typglaube einer Einheit in Beweismitteln niederschlägt. Die Parallelen zwischen Transaktionen und dem Eingehen von Bürgschaften sind dabei unübersehbar: In beiden Fällen geht eine Vertrauensentscheidung voraus. Auch bei Bürgschaftsbeziehungen ist nämlich abzuwägen, ob aus der Bindung Vor- oder Nachteile zu erwarten sind. Diese Vertrauensentscheidung führt zur Festlegung der Rahmenbedingungen. Hierfür wird bei Bürgschaftsbeziehungen bestimmt, welche Einheit der Partner ist und wie der Gültigkeitszeit-

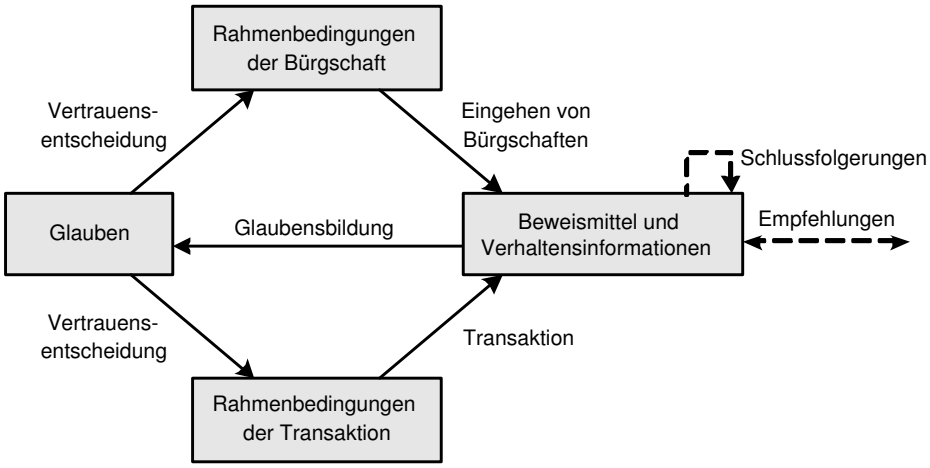

Abbildung 8.2: Erweiterung des Kreislaufs der verteilten Vertrauensbildung um soziale Beweismittel (datenzentrische Sicht)

raum zu gestalten ist. Abschließend findet sowohl bei Transaktionen als auch bei Bürgschaftsbeziehungen ein Austausch statt, der zusätzliche Beweismittel hervorbringt. Es zeigt sich also, dass wir durch den Einsatz von Bürgschaften sozusagen einen doppelten Kreislauf erhalten. Dabei hat jeder einzelne Kreislauf seine eigene Berechtigung. Der Austausch von Informationen, die von den menschlichen Prinzipalen benötigt werden, findet in den Transaktionen statt, während Bürgschaftsbeziehungen das soziale Gefüge explizit machen. Hierdurch wird die soziale Kontrolle weiter verschärft.

Lebenszyklus sozialer Beweismittel. Die beiden Sichten auf den Kreislauf der verteilten Vertrauensbildung geben einen Überblick, wie sich der Einsatz von sozialen Beweismitteln auf jede einzelne Einheit auswirkt. Im Folgenden stellen wir diesen beiden Sichten eine andere Sichtweise gegenüber, die vom Lebenszyklus eines einzelnen sozialen Beweismittels ausgeht. Dadurch wird das Zusammenspiel unterschiedlicher Einheiten verdeutlicht.

Abbildung 8.3 gibt eine abstrakte Sicht auf den Lebenszyklus. Sie wird in Abbildung 8.4 anhand eines Beispielablaufs zwischen drei Einheiten X, Y und Z illustriert. In einem Vorschritt machen die Einheiten X und Y positive Transaktionserfahrungen, so dass sie sich gegenseitig als hinreichend normativ ansehen. In den Folgeschritten führt dies zum Einsatz von sozialen Beweismitteln:

1. *Ausstellung:* Einheit X und Y gehen eine soziale Bindung ein, indem sie sich gegenseitig Bürgschaften ausstellen. In der Abbildung werden diese mit $B_X(Y)$ (Einheit X bürgt für Einheit Y) und $B_Y(X)$ bezeichnet.

2. *Weitergabe:* Einheit X bezieht die Bürgschaft $B_Y(X)$ der Einheit Y in seine Selbstempfehlung an Einheit Z ein. Dadurch kommt es zur Weitergabe des sozialen Beweismittels.

3. *Berücksichtigung:* Als Adressat der Selbstempfehlung empfängt Einheit Z die Bürgschaft. Sie nimmt von der Bürgschaftsbeziehung Kenntnis und passt ihren Typglauben entsprechend an.

4. *Löschung:* Nach Ablauf des Gültigkeitszeitraums der Bürgschaften $B_X(Y)$ und $B_Y(X)$ werden diese von allen Einheiten gelöscht, die sie bei sich lokal abgelegt haben. Insbesondere

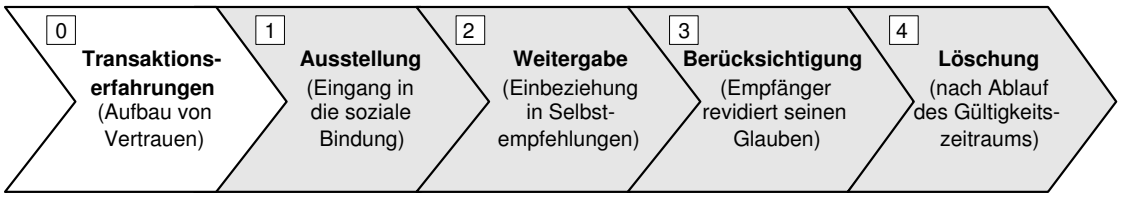

Abbildung 8.3: Der Lebenszyklus eines sozialen Beweismittels

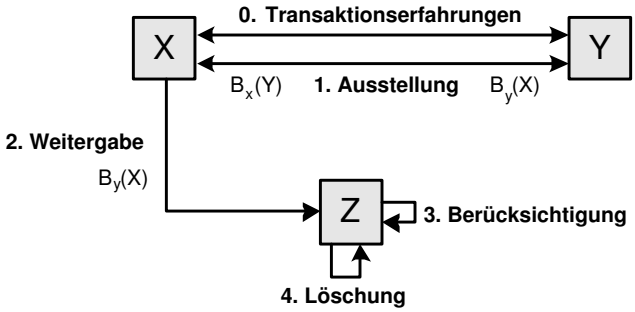

Abbildung 8.4: Illustration des Lebenszyklus sozialer Beweismittel anhand dreier Einheiten

löscht Einheit Z die Bürgschaft $B_Y(X)$ aus seiner Beweismittel-Basis und setzt ihren Typglauben entsprechend zurück

Die Illustration anhand dreier Einheiten ist insofern ein Beispiel, als die Phasen 2 und 3 des Lebenszyklus für die Bürgschaften $B_Y(X)$ und $B_X(Y)$ wiederholt durchlaufen werden können. Dies ist der Fall, wenn sich Einheit X beziehungsweise Einheit Y weiteren Einheiten gegenüber mit der jeweiligen Bürgschaft selbst empfiehlt.

8.2 Einsatz von Bürgschaften

Durch den Einsatz von Bürgschaften als soziale Beweismittel werden einige Erweiterungen des Entwurfs notwendig. Im Folgenden befassen wir uns damit,

- wie Bürgschaften ausgestellt werden (Abschnitt 8.2.1),
- wie sie Eingang in das Empfehlungssystem finden (Abschnitt 8.2.2) und
- wie sie von jeder Einheit lokal verwaltet werden (Abschnitt 8.2.3).

Zwei weitere Aspekte im Umgang mit Bürgschaften sind die Fragen, wie Bürgschaften in die Glaubensbildung einfließen und wann sie eingegangen werden. Diese Fragen werden separat in den Abschnitten 8.3 und 8.4 behandelt, da ihnen eine besondere Wichtigkeit zukommt.

8.2.1 Ausstellen von Bürgschaften als Transaktion

Im Folgenden gehen wir davon aus, dass zwei Einheiten bereit sind füreinander zu bürgen und sich bereits über die Gültigkeitsdauer der Bürgschaftsbeziehung abgesprochen haben. Dazu stellen wir uns die Frage, wie beide Einheiten diese Absprache durch das jeweilige Ausstellen einer Bürgschaft in einem sozialen Beweismittel festhalten.

Für den Austausch der Bürgschaften bietet sich das Sechs-Wege Transaktionsprotokoll aus Abschnitt 7.2.2 an. Dabei entspricht das jeweilige Ausstellen der Bürgschaft der Aktionsausführung im dritten und vierten Schritt des Protokolls. Durch den Austausch der Bürgschaften im Rahmen des Sechs-Wege Protokolls entsteht derselbe Schutz vor Betrugsverhalten wie in herkömmlichen Transaktionen, da der Austausch der Bürgschaften mit Verträgen und Quittungen dokumentiert wird. Verweigert eine der beiden Einheiten das Ausstellen der Bürgschaft, so ist sie von ihrem Partner im Rahmen des Empfehlungssystems angreifbar. Abgesehen von diesem Schutz gibt es eine Reihe weiterer Gründe, warum beim Austausch von sozialen Beweismitteln Betrugsverhalten nicht zu erwarten ist. Diese untersuchen wir im Folgenden.

Das Ausstellen einer Bürgschaft erfordert lediglich das Ausführen einer kryptographischen Operation. Im Vergleich zur Aktionsausführung in herkömmlichen Transaktionen wird damit weitaus weniger Zeit für den Austausch der Bürgschaften benötigt. Die schlägt sich in der Einschätzung der Wahrscheinlichkeit unbeabsichtigten Betrugsverhaltens nieder: Aufgrund der Kürze des Austausches sind Kommunikationsabbrüche besonders unwahrscheinlich. Gemäß der Revisionsvorschriften aus Abschnitt 6.3.3 führt damit Betrugsverhalten zu einer weitaus stärken Abwertung beim Transaktionspartner als in einer herkömmlichen Transaktion. Damit sind die Betrugskosten beim Austausch von Bürgschaften besonders hoch.

Auf der anderen Seite ist der Nutzen von Betrugsverhalten sehr gering. Angenommen eine Einheit X verweigert wider der Absprache das Ausstellen einer Bürgschaft für Einheit Y. Dies hat zur Folge, dass sich Einheit Y nicht mit der Bürgschaft der Einheit X selbst empfehlen kann. Einheit X verzichtet somit darauf, die Bürgschaft als eine Art Investition einzugehen. Dies ist aber zu ihrem Nachteil, wenn sich Einheit Y in der Folge anderen Einheiten gegenüber kooperativ verhält. In diesem Fall würde nämlich Einheit X für die Bürgschaft belohnt. Wir erhalten somit, dass sich das Verweigern des Bürgens für Einheit X nur dann lohnen kann, wenn von der Einheit Y in der Folge verstärkt Betrugsverhalten erwartet wird. In diesem Fall bringt jedoch die Bürgschaftsbeziehung mit der Einheit Y für Einheit X keinen Nutzen mit sich. Wenn sich Einheit X anderen Einheiten gegenüber selbst empfehlen will, wird sie nämlich darauf verzichten, Einheit Y als Bürgen aufzulisten. Dies liegt daran, dass das von ihr erwartete Betrugsverhalten in der Folge auch zu einer Abwertung der Einheit X führen wird. Somit erhalten wir insgesamt, dass Einheit X keinen Nutzen daraus ziehen kann, der Einheit Y die Bürgschaft vorzuenthalten.

8.2.2 Bürgschaften in Empfehlungen

Der Einsatz sozialer Beweismittel zielt darauf, den Einheiten die Möglichkeit zu aussagekräftigen Selbstempfehlungen zu geben. Dementsprechend sind Bürgschaften im Rahmen des Empfehlungssystems zu behandeln. Konkret bedeutet dies, dass Selbstempfehlungen nicht nur aus Quittungen sondern auch aus Bürgschaften bestehen können. Diese Erweiterung ändert nichts an den Vorschriften aus Abschnitt 7.3.2, die bestimmen, wann und wie Selbstempfehlungen ausgestellt werden. Lediglich der Bezug dieser Vorschriften ändert sich von Quittungen auf Bürgschaften. Für das Ausstellen von Selbstempfehlungen besteht weiterhin ein inhärenter Anreiz, da sich der Empfehler durch das Vorweisen seiner Bürgen als normativer darstellen kann.

Wie geht auf der anderen Seite eine Einheit bei Erhalt einer Selbstempfehlung vor? Nehmen wir die Situation aus Abbildung 8.4 an, in der Einheit Z die Selbstempfehlung der Einheit X und die darin enthaltene Bürgschaft der Einheit Y erhält. Daraus kann Einheit Z ableiten, dass Einheit X und Y eine Bürgschaftsbeziehung besitzen. Somit bringt es für Einheit Z keine neuen Erkenntnisse, wenn sie eine Selbstempfehlung der Einheit Y mit der Bürgschaft der Einheit X

erhält, da sie bereits über die Bürgschaftsbeziehung unterrichtet ist. In die Glaubensbildung geht daher immer nur das Wissen über eine Bürgschaftsbeziehung, nicht aber die einzelnen Bürgschaften ein. Eine Bürgschaftsbeziehung wird immer dann erkannt, wenn von einem der beiden Partner eine entsprechende Selbstempfehlung erhalten worden ist.

8.2.3 Verwaltung von Bürgschaften

Der Einsatz von Bürgschaften erfordert eine Erweiterung der Beweismittel- und Wissensverwaltung aus Abschnitt 7.5. Sie muss zusätzlich zu den folgenden Punkten in der Lage sein:

- *Ablage:* Die Bürgschaften, von denen die Einheit erfahren hat, werden in die Beweismittel-Basis abgelegt. Bei diesen Bürgschaften lässt sich differenzieren zwischen den beiden folgenden Arten: **(1)** Bürgschaften, bei denen die Einheit selbst der Gebürgte ist, werden im Zuge des Eingangs in die Bürgschaftsbeziehung vom Partner erhalten. **(2)** Von Bürgschaften, die die Bürgschaftsbeziehung zweier anderer Einheiten dokumentieren, wird durch den Empfang von Selbstempfehlungen erfahren.

- *Zugriff auf eigene Bürgschaftsbeziehungen:* Um Selbstempfehlungen zusammenstellen zu können, muss das Empfehlungssystem Zugang zu den Bürgschaften haben. Dabei sind die Bürgschaften erster Art von Interesse, da sie die eigenen Bürgschaftsbeziehungen belegen.

- *Kenntnisnahme vom Ablauf der eigenen Bürgschaftsbeziehungen:* Wird der Gültigkeitszeitraum einer eigenen Bürgschaftsbeziehung überschritten, so entfernt die Verwaltungskomponente die ungültig gewordene Bürgschaft aus der Beweismittel-Ablage. Um bei Bedarf die Bürgschaftsbeziehung zu erneuern, muss die Komponente zum Eingang von Bürgschaftsbeziehungen darüber benachrichtigt werden.

- *Zugriff auf fremde Bürgschaftsbeziehungen:* Bei der Glaubensbildung über eine Einheit ist von Interesse, mit welchen anderen Einheiten sie eine Bürgschaftsbeziehung besitzt. Zu diesem Zweck muss die Verwaltungskomponente in der Lage sein, die Bürgschaftsbeziehungen einer jeder Einheit anhand der bekannten Bürgschaften abzuleiten.

Diese Anforderungen lassen sich umsetzen, ohne dass die Architektur der Verwaltungskomponente aus Abschnitt 7.5.2 geändert werden müsste. Es sind lediglich weitere Regeln in das zugrunde liegende regelbasierte System hinzuzufügen. Dies lässt sich analog zu den Regeln für transaktionale Beweismittel durchführen.

8.3 Glaubensbildung mit Bürgschaften

Im vorhergehenden Abschnitt haben wir uns damit beschäftigt, wie Bürgschaftsbeziehungen entstehen und bekannt gemacht werden. In diesem Zusammenhang stellt sich die Frage, wie das Wissen um eine Bürgschaftsbeziehung in der Glaubensbildung berücksichtigt werden muss.

Diese Frage erörtern wir in diesem Abschnitt. Hierfür erweitern wir das Glaubensmodell in Abschnitt 8.3.1 um den sozialen Typglauben, der sich aus dem bisher behandelten individuellen Typglauben und der Kenntnis von Bürgschaftsbeziehungen zusammensetzt. Anschließend beschäftigen wir uns in Abschnitt 8.3.2 mit der Frage, wie der Zusammenhang zwischen individuellem und sozialem Typglauben bei der Durchführung von Glaubensrevisionen gewahrt werden kann.

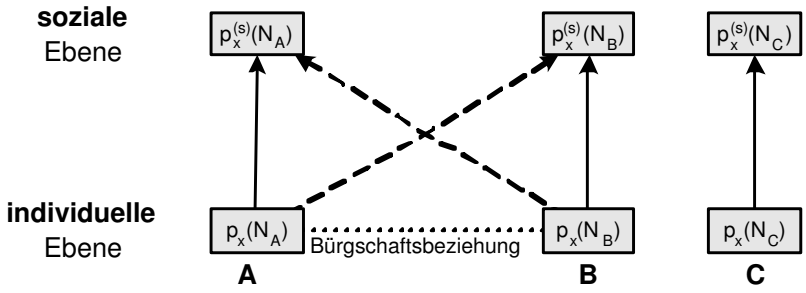

Abbildung 8.5: Zusammenhang zwischen der individuellen und sozialen Ebene der Glaubensbildung

8.3.1 Sozialer Typglaube

Im Folgenden wird eine weitere Ebene des Typglaubens, der soziale Typglaube, vorgestellt, in die das Wissen um Bürgschaftsbeziehungen mit einfließt. Hierfür wird zunächst besprochen, was unter sozialem Typglauben zu verstehen ist und wie er sich vom bisher betrachteten individuellen Typglauben unterscheidet. Anschließend werden die Berechnungsvorschriften für die Quantifizierung des sozialen Typglaubens vorgestellt. Die Wirkungsweise dieser Vorschriften wird abschließend anhand eines Beispiels illustriert.

Ebenen des Typglaubens. In den Kapiteln 6 und 7 sind wir von einem Typglauben ausgegangen, der keine sozialen Bindungen kennt. Diesen Typglauben nennen wir *individuell*, da er lediglich vom Verhalten der jeweiligen Einheit und nicht etwa vom Typglauben über andere Einheiten beeinflusst wird. Somit ist der individuelle Typglaube nicht in der Lage, das Wissen um Bürgschaftsbeziehungen einzubeziehen.

Daher führen wir eine weitere Ebene des Typglaubens ein. Dieser *soziale Typglaube* ergibt sich aus dem individuellen Typglauben einerseits und dem Wissen über Bürgschaftsbeziehungen andererseits. Von der Notation her kennzeichnen wir ihn mit einem hochgestellten (s) für sozial. Der soziale Typglaube der Einheit X über eine Einheit Y wird somit mit $p_X^{(s)}(N_Y)$ bezeichnet.

Wie ist der Zusammenhang zwischen dem individuellen Typglauben $p_X(N_Y)$ und dem sozialen Typglauben $p_X^{(s)}(N_Y)$? Besitzt Einheit Y keine Bürgen, so gibt es laut Definition keinen Unterschied zwischen diesen beiden Ebenen des Typglaubens, das heißt $p_X(N_Y) = p_X^{(s)}(N_Y)$. Für den Fall, dass Einheit Y Bürgen besitzt, muss die Einheit X ihren individuellen Typglauben über diese Bürgen einfließen lassen.

Dieser Sachverhalt wird in Abbildung 8.5 verdeutlicht. Es ist eine schematische Darstellung des Typglaubens der Einheit X über die drei Einheiten A, B und C. In der Abbildung wird angenommen, dass Einheit X durch eine entsprechende Selbstempfehlung davon unterrichtet worden ist, dass eine Bürgschaftsbeziehung zwischen den Einheiten A und B besteht. Damit hängt der soziale Typglaube über Einheit A vom individuellen Typglauben sowohl über Einheit A (durchgezogener Pfeil) als auch Einheit B (gestrichelter Pfeil) ab. Dies ist sinnvoll, da der Glaube über die Normativität der Einheit B einen Aufschluss über den Typ der Einheit A gibt und daher zu berücksichtigen ist. Gleiches gilt aufgrund der Symmetrie von Bürgschaftsbeziehungen für den sozialen Typglauben über Einheit B. Hingegen stimmen der individuelle und soziale Typglaube über Einheit C überein, da keine Bürgschaftsbeziehungen von ihr bekannt sind.

Weil Bürgschaften zeitlich begrenzt sind und damit ungültig werden können, sind Bürgschafts-

beziehungen ihrer Natur nach transient. Daher bietet es sich an, dass der soziale Typglaube über eine Einheit bei Bedarf aus dem individuellen Typglauben abgeleitet wird. Dies hat zum Vorteil, dass die Berücksichtigung und Löschung von Bürgschaftsbeziehungen nicht einen erheblichen Aufwand mit sich ziehen[4]. Wenn der soziale Typglaube nicht gespeichert sondern nur bei Bedarf abgeleitet wird, stellt die soziale Ebene des Typglaubens eine Sicht auf die individuelle Ebene des Typglaubens dar. Es reicht somit aus, dass nur der individuelle Typglaube materialisiert, das heißt explizit gespeichert, wird. Wenn die Odds-Darstellung zum Einsatz kommt, hält sich der Aufwand für die bedarfsorientierte Berechnung des sozialen Typglaubens in Grenzen, wie die Berechnungsvorschriften des nächsten Paragraphs zeigen.

Der soziale Typglaube findet überall dort Anwendung, wo bisher der Einsatz des individuellen Typglaubens vorgesehen war. Dies betrifft also zwei Gebiete:

- *Vertrauensentscheidungen:* Auch aus dem sozialen Typglauben lässt sich eine probabilistische Einschätzung des Verhaltens möglicher Transaktionspartner durchführen. Hierfür ist dies TIB-Modell lediglich auf den sozialen und nicht auf den individuellen Typglauben anzuwenden.

- *Glaubensrevision:* Wird von der Beweismittel- und Wissensverwaltung ein neues Ereignis ausgelöst, so kommen die bisher hergeleiteten Revisionsvorschriften zum Einsatz. Diese werden direkt auf den sozialen Typglauben angewandt. Hierbei ist das Problem zu lösen, dass der soziale Typglaube lediglich eine Sicht auf den individuellen Typglauben ist und somit nicht materialisiert ist. Abschnitt 8.3.2 wird sich mit diesem Problem befassen und es lösen.

Quantifizierung des sozialen Typglaubens. In einer Bürgschaftsbeziehung bestätigen sich die beiden Partner gegenseitig, dass sie denselben Typ besitzen. Wenn Einheit A und B, wie in Abbildung 8.5 dargestellt, eine Bürgschaftsbeziehung besitzen, so bedeutet dies für den Typglauben der Einheit X über Einheit A das Folgende: Entweder sind sowohl Einheit A als auch B normativ, oder sie sind beide strategisch. Damit schließen wir zunächst aus, dass eine normative Einheit irrtümlicherweise für eine strategische Einheit bürgen könnte. Damit ergibt sich der soziale Typglaube über Einheit A wie folgt:

$$\begin{aligned} p_X^{(s)}(N_A) &= p_X^{(s)}(N_A \wedge N_B) + p_X^{(s)}(N_A \wedge S_B) = p_X^{(s)}(N_A \wedge N_B) + 0 \\ &= \frac{p_X(N_A \wedge N_B)}{p_X(N_A \wedge N_B) + p_X(S_A \wedge S_B)} = \frac{p_X(N_A) \cdot p_X(N_B)}{p_X(N_A) \cdot p_X(N_B) + p_X(S_A) \cdot p_X(S_B)} \end{aligned} \quad (8.1)$$

Diese Berechnungsvorschrift vereinfacht sich, wenn wir den sozialen Typglauben in der Odds-Darstellung bestimmen:

$$\begin{aligned} \hat{p}_X^{(s)}(N_A) &= \frac{1}{p_X^{(s)}(N_A)} - 1 = \frac{p_X(N_A) \cdot p_X(N_B) + p_X(S_A) \cdot p_X(S_B)}{p_X(N_A) \cdot p_X(N_B)} - 1 \\ &= 1 + \frac{p_X(S_A) \cdot p_X(S_B)}{p_X(S_A) \cdot p_X(S_B)} - 1 = \frac{p_X(S_A)}{p_X(N_A)} \cdot \frac{p_X(S_B)}{p_X(N_B)} = \hat{p}_X(N_A) \cdot \hat{p}_X(N_B) \end{aligned} \quad (8.2)$$

Somit berechnet sich der soziale Typglaube über Einheit A in der Odds-Darstellung einfach als das Produkt des individuellen Typglaubens über Einheit A und desjenigen über Einheit B. Dieses

[4]Gemäß der Revisionsvorschriften aus Abschnitt 8.3.2 würde ein Aufwand verursacht werden, der sich quadratisch zur Zahl der Bürgschaftsbeziehungen der beteiligten Einheiten verhält.

Ergebnis lässt sich wie folgt verallgemeinern: Sei $\beta_X(Y)$ die Menge der Bürgen der Einheit Y, von denen Einheit X weiß. Dann lässt sich der soziale Typglaube über Einheit Y gemäß der folgenden Formel berechnen:

$$\hat{p}_X^{(s)}(N_Y) = \hat{p}_X(N_Y) \cdot \prod_{E \in \beta_X(Y)} \hat{p}_X(N_E) \tag{8.3}$$

Aus dieser Berechnungsvorschrift lassen sich die Eigenschaften des sozialen Typglaubens erkennen. Es handelt sich bei ihnen um die Folgenden:

- *Neutrale Bürgen:* Angenommen, der individuellen Typglaube über Einheit B beträgt $p_X(N_B) = 50\%$, so ergibt sich $\hat{p}_X(N_B) = 1$ und somit $p_X^{(s)}(N_A) = p_X(N_A)$. Dies bedeutet, dass sich in diesem Fall die Bürgschaft der Einheit B auf den sozialen Typglauben über die Einheit A nicht auswirkt. Insofern handelt es sich bei der Einheit B um einen neutralen Bürgen der Einheit A. Dieses Ergebnis ist probabilistisch sinnvoll, da ein Typglaube von 50% aussagt, dass Einheit B ebenso wahrscheinlich normativ wie strategisch ist. Die Wahrscheinlichkeit, dass Einheit A normativ ist, verändert sich somit nicht, wenn wir erfahren, dass sowohl Einheit A als auch Einheit B denselben Typ besitzen.

- *Normative oder strategische Bürgen:* Ist unter den Bürgen eine Einheit, die als sicher normativ ($p_X(N_B) = 1$) oder sicher strategisch ($p_X(N_B) = 0$) erscheint, so ergibt sich ein sicherer Glaube über den Typ des Gebürgten. Diese Feststellung leitet sich in der Formel daraus ab, dass in diesem Fall die Odds-Darstellung des Typglaubens über den Bürgen $\hat{p}_X(N_B) = 0$ beziehungsweise $\hat{p}_X(N_B) = \infty$ ist und dieser Faktor bei der Berechnung des sozialen Typglaubens über Einheit A dominiert. Auch dieses Ergebnis ist probabilistisch sinnvoll, da aufgrund der angenommenen Übereinstimmung des Typs eine Sicherheit über den Typ des Bürgendens auf den Typ des Gebürgten abfärbt.

- *Unmögliche Bürgschaften:* Was passiert allerdings, wenn eine Einheit zwei Bürgen besitzt, wobei der Eine als sicher normativ und der andere als sicher strategisch erscheint? In diesem Fall erhalten wir, dass die Berechnungsvorschrift aufgrund der Multiplikation von null und unendlich undefiniert ist. Es ist somit Sache des Entwerfers die Definition der Vorschrift sinnvoll zu erweitern. Gemäß der Vorschriften zur Glaubensbildung kann ein sicherer Glaube nur aus Typinformationen abgeleitet worden sein. Eine Möglichkeit, dieses Problem zu lösen, besteht also darin, erst gar nicht externe Typinformationen zuzulassen. Bei intern abgeleiteten Typinformationen handelt es sich nämlich ausschließlich um Typbeweisen über eine inkonsistente Einheit, die somit als sicher strategisch erscheint. Durch Ausschluss externer Typinformationen können also sich widersprechende Bürgschaften vermieden werden[5].

Abgeschwächte Berücksichtigung von Bürgschaften im sozialen Typglauben. Bei der Quantifizierung des sozialen Typglaubens sind wir bisher davon ausgegangen, dass die Aussage der Bürgschaftsbeziehung, die Übereinstimmung der Typen der beiden Partner, zutrifft. Allerdings haben wir schon zu einem früheren Zeitpunkt festgestellt, dass Einheiten sich beim Eingehen von Bürgschaftsbeziehungen irren können. So ist es zum Beispiel nicht auszuschließen, dass eine normative Einheit für eine strategische Einheit bürgt, da diese ihr bisher normativ erschienen ist.

[5]Es gibt eine ganze Reihe von weiteren Möglichkeiten, mit sich widersprechenden Bürgschaften umzugehen. Eine besteht darin, Typinformationen nicht als absolut sicher anzusehen. Konkret bedeutet dies, dass die Richtigkeit einer Typinformation lediglich mit der Wahrscheinlichkeit $1 - \epsilon$ (mit einem kleinem ϵ) angenommen wird. Diesen Ansatz betrachten wir aber im Folgenden nicht weiter, da wir insbesondere im Evaluationsteil dieser Arbeit keine externen Quellen von Typinformationen annehmen und es somit zu keinerlei Widerspruch kommen kann.

Wir benötigen demnach eine Möglichkeit, den Einfluss der Bürgschaften auf die Berechnung des sozialen Typglaubens abzuschwächen.

Um dieses Ziel zu erreichen, führen wir einen Parameter λ für die Berechnungsvorschrift des sozialen Typglaubens ein. Dieser bewegt sich im Bereich $[0, 1]$. Dabei gibt ein Wert von eins an, dass Bürgschaften im Sinne der Gleichung 8.3 voll eingerechnet werden. Hingegen bedeutet ein Wert von null, dass Bürgschaften nicht berücksichtigt werden und der soziale Typglaube somit mit dem individuellen Typglauben übereinstimmt. Im Bereich zwischen null und eins erfolgt eine graduell stärker werdende Einbeziehung der Bürgschaften. Diese Anforderungen werden von folgender Verfeinerung der Gleichung 8.3 erfüllt:

$$\hat{p}_X^{(s)}(N_Y) = \hat{p}_X(N_Y) \cdot [\prod_{E \in \beta_X(Y)} \hat{p}_X(N_E)]^{\lambda} \tag{8.4}$$

Durch die Wahl einer passenden Größe des Parameters λ ist der Entwerfer in der Lage, seine Vorstellung von Bürgschaftsbeziehungen dem jeweiligen Anwendungsgebiet anzupassen. Hierbei ist die Abhängigkeit von den Vorschriften zum Eingehen von Bürgschaftsbeziehungen, die in Abschnitt 8.4.1 eingeführt werden, zu beachten: Gehen Einheiten nur dann Bürgschaften ein, wenn sie sich gegenseitig mit hoher Wahrscheinlichkeit als normativ ansehen, ist ein Wert von λ nahe eins sinngemäß, da dann ein irrtümliches Bürgen eher unwahrscheinlich ist. Wenn allerdings Einheiten bereits bei einem geringen Typglauben Bürgschaftsbeziehungen eingehen, sollte der Parameter λ nicht zu hoch sein. Ein Beispiel für die Festlegung von λ und der Parametrisierung der Vorschrift zum Eingehen von Bürgschaftsbeziehungen findet sich in Abschnitt A.2.1 des Anhangs.

Illustration an einem Beispiel. Wir erweitern die Situation aus Abbildung 8.5, indem Einheit X drei weitere Einheiten D, E und F kennt. Dabei gibt es eine Bürgschaftsbeziehung zwischen Einheit D und E beziehungsweise zwischen Einheit E und F. Das sich ergebende Netz aus Bürgschaftsbeziehungen zeigt Abbildung 8.6.

Tabelle 8.1 gibt ein Beispiel für die Berechnung des sozialen Typglaubens aus der Sicht der Einheit X. Hierfür wird der individuelle Typglaube über die sechs Einheiten vorgegeben und die Berechnungsreihenfolge dargestellt. Dabei gehen wir von einem Parameter $\lambda = 0.5$ aus. Das bedeutet, dass der individuelle Typglaube über Bürgen halb so viel Gewicht erfährt wie der individuelle Typglaube über die eigentliche Einheit. Vor der Berechnung des sozialen Typglaubens in Odds-Darstellung ist in der dritten Zeile der Faktor dargestellt, der den Einfluss der Bürgen erfasst. Aus der Tabelle sind folgende Punkte hervorzuheben:

- Durch die Bürgschaftsbeziehung mit Einheit B ist der soziale Typglaube über Einheit A höher als derjenige über Einheit C, obwohl es sich beim individuellen Typglauben genau umgekehrt verhält. Dies ergibt sich daraus, dass Einheit B ein ziemlich normativ erscheinender Bürge ist.

- Einheit E besitzt je einen Bürgen, der strategisch (Einheit D) beziehungsweise normativ (Einheit F) erscheint. Im Ergebnis gleicht sich der Einfluss der Bürgen damit fast aus. Hätte Einheit E auf die Bürgschaftsbeziehung mit Einheit D verzichtet, so würde sie ähnlich wie Einheit A aufgewertet.

- Wie in Abschnitt 8.1.1 gezeigt, lassen sich keine transitiven Schlüsse aus Bürgschaftsbeziehungen ziehen. Im Beispiel bedeutet dies, dass Einheit D und F keine Bürgschaftsbeziehung besitzen, obwohl sie sich mit derselben Einheit E gebunden haben. In den Berechnungsvorschriften kommt dies zum Ausdruck, indem der individuelle Typglaube über Einheit D

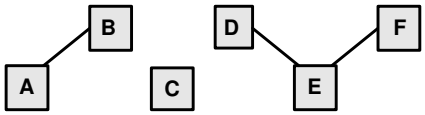

Abbildung 8.6: Beispiel für ein Netz aus Bürgschaftsbeziehungen

Tabelle 8.1: Beispielhafte Berechnung des sozialen Typglaubens

Einheit i	**A**	**B**	**C**	**D**	**E**	**F**
$p_X(N_i)$	**60%**	**80%**	**65%**	**20%**	**55%**	**70%**
$\hat{p}_X(N_i)$	0,667	0,250	0,538	4,000	0,818	0,429
$\prod_{j\in\beta_X(i)} \hat{p}_X(N_j)$	0,250	0,667	1,000	0,818	1,716	0,818
$\hat{p}_X^{(s)}(N_i)$	0,333	0,204	0,538	3,618	1,072	0,388
$p_X^{(s)}(N_i)$	**75%**	**83%**	**65%**	**22%**	**48%**	**72%**

und F nicht gegenseitig in die Berechnung des sozialen Typglaubens über sie eingeht. Damit ist Einheit F vor der (unter Umständen) irrtümlichen Bürgschaftsbeziehung der Einheit E mit Einheit D geschützt. Auf der anderen Seite erhält Einheit D nicht die Vorteile aus einer Bürgschaft der Einheit F.

8.3.2 Glaubensrevision

Grundlage für die Durchführung von Glaubensrevisionen ist der im vorhergehenden Abschnitt eingeführte soziale Typglaube. Dieser ist lediglich eine Sicht auf den individuellen Typglauben. Daher kann der soziale Typglaube de facto nur dann revidiert werden, wenn sich auch der zugrunde liegende individuelle Typglaube ändert. Dabei fordern wir, dass es von außen den Anschein hat, als ob die anzuwendende Revisionsvorschrift direkt auf dem sozialen Typglauben angewendet worden wäre.

Mit der Umsetzung dieser Anforderung befassen wir uns in diesem Abschnitt. Dabei entwickeln wir zunächst ein Modell, das diese Anforderung erfüllt und verfeinern es anschließend mit der Möglichkeit zur Parametrisierung. Die Wirkungsweise der Revisionsvorschriften wird abschließend anhand eines Beispiels illustriert.

Modell zur Revision des sozialen Typglaubens. Abbildung 8.7 erweitert das Beispiel aus Abbildung 8.5, das die Glaubensbildung der Einheit X über die Einheiten A, B und C zeigt. Dabei wird davon ausgegangen, dass Einheit X von der Bürgschaftsbeziehung zwischen den Einheiten A und B unterrichtet ist. Anhand dieses Beispiels wird ersichtlich, wie ein Modell zur Revision des sozialen Typglaubens zu entwickeln ist. Anstoßpunkte sind hierbei Ereignisse, die die Revision des Typglaubens über Einheit A und C nahe legen:

- *Revision des Glaubens über Einheit C:* Sei r der Revisionsfaktor, mit dem der Typglauben über Einheit C laut Revisionsvorschrift zur revidieren ist. Da die Revision auf dem sozialen Typglauben durchzuführen ist, erhalten wir den posterioren Typglauben durch Multiplikation des Revisionsfaktors mit dem prioren Typglauben, also $r \cdot \hat{p}_X^{(s)}(N_C)$. Das Problem ist, dass nicht am sozialen Typglauben direkt eine Änderung durchgeführt werden kann, da dieser

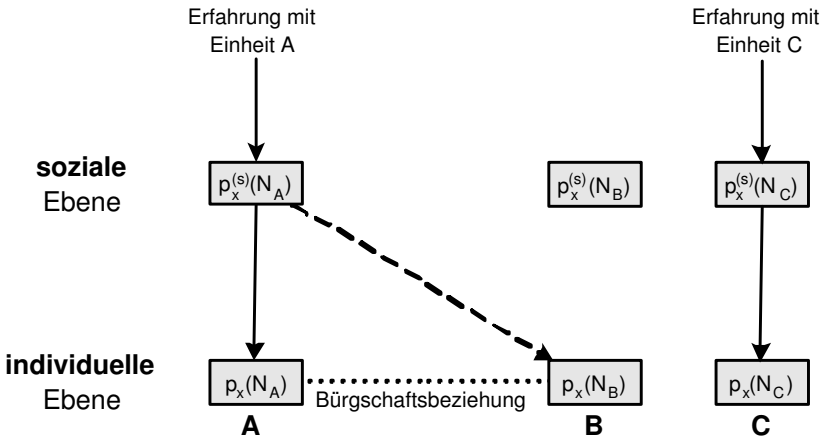

Abbildung 8.7: Revision des sozialen Typglaubens und Rückwirkung auf den individuellen Typglauben

sich immer bei Bedarf aus dem individuellen Typglauben ableitet. Es ist also vielmehr der individuelle Typglauben so anzupassen, dass sich für den posterioren sozialen Typglaube tatsächlich $r \cdot \hat{p}_X^{(s)}(N_C)$ ergibt. Da Einheit C keine soziale Bindung besitzt, ist diese Aufgabe jedoch einfach zu lösen: Der individuelle und soziale Typglaube stimmen überein. Der Revisionsfaktor r kann somit direkt auf den individuellen Typglauben angewendet werden. Dadurch erhalten wir das gewünschte Ergebnis für den posterioren sozialen Typglauben.

- *Revision des Glaubens über Einheit A:* Schwieriger ist die Situation, wenn eine Revision des Glaubens über Einheit A ansteht. Dies liegt an der Bürgschaftsbeziehung zwischen Einheit A und B. In den sozialen Typglauben über Einheit A geht daher auch der individuelle Typglaube über die Einheit B ein. Dadurch stellt sich die Frage, wie der individuelle Typglaube über diese beiden Einheiten angepasst werden muss. Der Revisionsfaktor r muss dabei aufgeteilt werden. Wie dies geschieht zeigen wir zunächst für einen Spezialfall: Nehmen wir an, dass Bürgschaftsbeziehungen voll berücksichtigt werden, für den Parameter also $\lambda = 1$ gilt. In diesem Fall ergibt sich der soziale Typglaube über Einheit A gemäß der Gleichung 8.2 als einfaches Produkt. Der Revisionsfaktor lässt sich somit auf die folgende Weise aufteilen (das Revisionsereignis wird mit R bezeichnet):

$$\begin{aligned} \hat{p}_X^{(s)}(N_A|R) &\stackrel{!}{=} r \cdot \hat{p}_X^{(s)}(N_A) \\ \hat{p}_X(N_A|R) \cdot \hat{p}_X(N_B|R) &= r \cdot \hat{p}_X(N_A) \cdot \hat{p}_X(N_B) \\ &= (r^{\frac{1}{2}} \cdot \hat{p}_X(N_A)) \cdot (r^{\frac{1}{2}} \cdot \hat{p}_X(N_B)) \end{aligned} \tag{8.5}$$

Damit ergibt sich für die Revision des individuellen Typglaubens über Einheit A und B jeweils der Revisionsfaktor $r^{\frac{1}{2}}$. Diese Aufteilung des Revisionsfaktors in gleichen Teilen ist sinnvoll, wenn wir annehmen, dass die Aussage der Bürgschaftsbeziehung, die Typgleichheit der beiden Partner, zutrifft. Außerdem wird durch die Revision des individuellen Glaubens über Einheit B der Gedanke umgesetzt, dass eine Bürgschaft eine Investition darstellt. Dies wird in der Abbildung durch den gestrichelten Pfeil visualisiert.

Wie lässt sich die Revision des sozialen Typglaubens über Einheit A verallgemeinern? Nehmen wir wie bereits in Abschnitt 8.3.1 an, dass $\beta_X(Y)$ die Menge der Einheiten ist, die Einheit X

als Bürgen der Einheit Y kennt. Wenn wir an der gleichen Aufteilung des Revisionsfaktors r festhalten, ergibt sich für den Fall $\lambda = 1$, dass der individuelle Typglaube über Einheit Y und über jeden ihrer Bürgen mit dem Faktor $r^{(1+|\beta_X(Y)|)^{-1}}$ revidiert werden muss. Dieses Ergebnis lässt sich auch für allgemeine Parameter λ verallgemeinern. Ausgangspunkt hierfür ist die Forderung, dass es von außen so aussieht, als ob der Faktor der Revision R direkt auf den sozialen Typglauben über die Einheit Y angewendet wird:

$$\begin{aligned} p_X^{(s)}(N_Y|R) &\stackrel{!}{=} r \cdot p_X^{(s)}(N_Y) \\ \hat{p}_X(N_Y|R) \cdot [\prod_{E \in \beta_X(Y)} \hat{p}_X(N_E|R)]^\lambda &= r \cdot \hat{p}_X(N_Y) \cdot [\prod_{E \in \beta_X(Y)} \hat{p}_X(N_E)]^\lambda \\ &= (\rho \cdot \hat{p}_X(N_Y)) \cdot [\prod_{E \in \beta_X(Y)} \rho \cdot \hat{p}_X(N_E)]^\lambda \end{aligned} \tag{8.6}$$

Dabei ist der Revisionsfaktor ρ, der auf den individuellen Typglauben über Einheit Y und ihre Bürgen anzuwenden ist, wie folgt abzuleiten:

$$r \stackrel{!}{=} \rho \cdot (\rho^{|\beta_X(Y)|})^\lambda = \rho^{1+|\beta_X(Y)| \cdot \lambda} \quad \Longrightarrow \quad \rho = r^{(1+|\beta_X(Y)| \cdot \lambda)^{-1}} \tag{8.7}$$

Für den Spezialfall $\lambda = 1$ wird somit der Faktor $r^{(1+|\beta_X(Y)|)^{-1}}$ bestätigt. Erhalten Bürgschaften hingegen kein Gewicht ($\lambda = 0$), so ergibt sich $\rho = r$. Damit wird der individuelle Typglaube über Einheit Y mit dem vollen Faktor revidiert.

Gemäß dieser Revisionsvorschrift des individuellen Typglaubens werden die Bürgen einer Einheit herangezogen, wenn über diese Einheit neue Informationen vorliegen. Dies ist sinnvoll, da dadurch für die Bürgschaftsbeziehungen der Aspekt der Investition durchgesetzt wird: Wenn sich die Festlegung des Bürgen, dass der Gebürgte normativ ist, in Transaktionen mit dem Gebürgten bewahrheitet (beziehungsweise sich als falsch herausstellt), muss auch der Bürge normativer (beziehungsweise strategischer) als zuvor wahrgenommen werden. Ebendies wird durch das entwickelte Modell zur Aufteilung des Revisionsfaktors erreicht.

Verfeinerung des Modells. Bisher sind wir davon ausgegangen, dass der Revisionsfaktor in gleichen Teilen auf die Einheit und ihre Bürgen aufgeteilt wird. In Analogie zu Abschnitt 8.3.1 müssen wir uns jedoch fragen, ob die Aussage der Bürgschaftsbeziehung als wahr angenommen werden kann. Da auch normative Einheiten mit strategischen Einheiten Bürgschaftsbeziehungen haben könnten, sollte der individuelle Typglaube über die Einheit stärker revidiert werden als der über ihre Bürgen. Somit benötigen wir einen weiteren Parameter ν, mit dem wir eine ungleiche Aufteilung des Revisionsfaktors vornehmen können. Dabei bedeutet $\nu = 1$, dass der Revisionsfaktor gleich aufgeteilt wird. Dies führt zu einer Glaubensrevision entsprechend Gleichung 8.6. Hingegen besagt $\nu = 0$, dass nur die Einheit, deren sozialer Typglaube zu revidieren ist, zur Revision des individuellen Typglaubens heranzuziehen ist.

Um diese Parametrisierung des Modells zu erreichen, müssen wir Gleichung 8.7 verfeinern. Dabei bezieht sich ρ nunmehr auf den Revisionsfaktor, der auf den individuellen Typglauben über Einheit Y angewandt wird. Für die Bürgen der Einheit Y wird der Revisionsfaktor ρ^ν benutzt. Dadurch erreichen wir unser Ziel, dass die Bürgen bei $\nu = 1$ in gleichen Teilen und bei $\nu = 0$ gar nicht in die Revision einbezogen werden. Der Faktor ρ bestimmt sich daher wie folgt aus r:

$$r \stackrel{!}{=} \rho \cdot ((\rho^\nu)^{|\beta_X(Y)|})^\lambda = \rho^{1+\nu \cdot |\beta_X(Y)| \cdot \lambda} \quad \Longrightarrow \quad \rho = r^{(1+|\beta_X(Y)| \cdot \nu \cdot \lambda)^{-1}} \tag{8.8}$$

Aus dieser Formel lassen sich einige Schlüsse für die Revision des sozialen Typglaubens ziehen:

- *Abhängigkeit von der Zahl der Bürgen:* Je mehr Bürgen eine Einheit besitzt, desto weniger wird der individuelle Typglaube über sie beeinflusst. In der Formel kommt dies zum Ausdruck, da sich ρ bei steigendem $|\beta_X(Y)|$ zum neutralen Wert eins hin bewegt. Dies liegt am Folgenden: Je mehr Bürgen eine Einheit hat, desto breiter wird die Revision gestreut und desto weniger wird in der Folge der individuelle Typglaube über sie verändert. Durch das Eingehen von Bürgschaftsbeziehungen kann sich also eine Einheit gegen das Risiko von eigenem unbeabsichtigtem Betrugsverhalten versichern. In Analogie zu menschlichen Gesellschaften erhöhen soziale Bindungen somit die Sicherheit, die eine Einheit genießt.

- *Abhängigkeit von der Gewichtung der Bürgen:* Die Bürgschaftsbeziehungen werden in zweierlei Hinsicht gewichtet. **(1)** Der Parameter λ gibt an, wie stark der individuelle Typglaube der Bürgen in die Berechnung des sozialen Typglaubens einfließt. Je stärker dieses Gewicht ist, desto neutraler fällt der Revisionsfaktor ρ aus. Dies liegt daran, dass sich die Revision des individuellen Typglaubens über die Bürgen stärker auf den sozialen Typglauben über die Einheit Y auswirkt. **(2)** Der Parameter ν bestimmt die Höhe des Revisionsfaktors für Einheit Y im Vergleich zu ihren Bürgen. Je größer das Gewicht der Einheit Y bei der Revision, desto stärker verändert sich der individuelle Typglaube über sie. In der Gleichung zeigt sich dies dadurch, dass sich der Revisionsfaktor ρ für kleine ν dem Faktor r annähert. Umgekehrt verhält es sich bei den Bürgen. Der individuelle Typglaube über sie wird genau dann besonders stark angepasst, wenn ihnen ein großes Gewicht (also ν nahe eins) bei der Revision zukommt.

Illustration an einem Beispiel. Durch eine Erweiterung des Beispiels aus Abbildung 8.6 lässt sich das Modell für die Revision des sozialen Typglaubens illustrieren. Wir gehen davon aus, dass aus der Sicht der Einheit X folgende Ereignisse stattfinden[6]:

1. Mit Einheit A wird eine positive Transaktionserfahrung gemacht. Der Revisionsfaktor r_C, der sich daraus ergibt, ist $0,5$.

2. In einer Transaktion mit Einheit D nimmt Einheit X Betrugsverhalten wahr. Dies führt zum Revisionsfaktor $r_D = 5,5$.

3. Einheit C wird von der Einheit B negativ empfohlen. Der Revisionsfaktor r_K ist abhängig vom Typglauben über Einheit B zu diesem Zeitpunkt. Er kann also erst nach der Einbeziehung der ersten beiden Ereignisse ausgerechnet werden.

4. Die Bürgschaftsbeziehung zwischen Einheit D und E wird ungültig. Sie wird auch nicht wieder erneuert.

Im Folgenden gehen wir diese vier Ereignisse Schritt für Schritt durch und untersuchen, welche Auswirkungen sich daraus für den Typglauben über die sechs Einheiten ergeben. Der jeweilige sich ergebende individuelle und soziale Typglaube über jede Einheit ist in Tabelle 8.2 zusammengefasst. Verändert sich der Typglaube über eine Einheit in einem Schritt, so wird dies der Übersichtlichkeit wegen fett hervorgehoben. In der Tabelle gehen wir von einem Parameter $\nu = 0,5$ aus. Das bedeutet, dass der individuelle Typglauben über die Bürgen nur halb so stark revidiert wird wie derjenige über die eigentliche Einheit. Die Schritte der Tabelle erklären sich wie folgt:

[6] Die Revisionsfaktoren ergeben sich für den Fall, dass für den jeweiligen Transaktionskontext $p_n(\gamma) = 50\%$ und $p_u(\gamma) = 5,5\%$ gilt.

Tabelle 8.2: Beispielhafte Anwendung des Modells zur Revision des sozialen Typglaubens

Schritt	**Einheit** i	**A**	**B**	**C**	**D**	**E**	**F**
0.	$p_X(N_i)$	60%	80%	65%	20%	55%	70%
(anfangs)	$p_X^{(s)}(N_i)$	75%	83%	65%	22%	48%	72%
1.	$p_X(N_i)$	**72%**	**84%**	65%	20%	55%	70%
	$p_X^{(s)}(N_i)$	**86%**	**89%**	65%	22%	48%	72%
2.	$p_X(N_i)$	72%	84%	65%	**6%**	**38%**	70%
	$p_X^{(s)}(N_i)$	86%	89%	65%	**5%**	**19%**	**65%**
3.	$p_X(N_i)$	72%	84%	**41%**	6%	38%	70%
	$p_X^{(s)}(N_i)$	86%	89%	**41%**	5%	19%	65%
4.	$p_X(N_i)$	72%	84%	41%	6%	38%	70%
	$p_X^{(s)}(N_i)$	86%	89%	41%	**6%**	**48%**	65%

1. Für $p_X^{(s)}(N_A|C)$ ergibt sich unter Anwendung des Revisionsfaktors $r_C = 0,5$ der Wert 86%. Für die Revision des individuellen Typglauben über Einheit A und B muss der Revisionsfaktor aufgeteilt werden. Gemäß der Formel 8.8 erhalten wir als Revisionsfaktor für Einheit A den Wert $\rho = 0,5^{1,25^{-1}} = 0,574$ und für Einheit B den Wert $\rho^{0,5} = 0,758$. Dies führt zu einem individuellen Typglauben über Einheit A und B von 72% beziehungsweise 84%. Wir können aus diesen beiden Werten zur Probe den posterioren sozialen Typglauben über Einheit A berechnen und werden dabei mit dem Wert 86% bestätigt. Der veränderte individuelle Typglaube über die Einheiten A und B führt zu einer Anpassung des sozialen Typglaubens über Einheit B. Dieser beträgt nunmehr 89%. Für Einheit B hat sich die Bürgschaftsbeziehung mit Einheit A also gelohnt, da diese sich kooperativ verhalten hat.

   Dieses erste Ereignis wirkt sich nicht auf den Typglauben über die anderen vier Einheiten aus, da sie keine Bürgschaftsbeziehung mit Einheit A haben.

2. Der Betrug der Einheit D führt zu einer analogen Glaubensrevision. Der soziale Typglaube über Einheit D vermindert sich auf 5%. Der Revisionsfaktor r_D wird aufgeteilt durch $\rho = 3,910$ und $\rho^{0,5} = 1,978$ und bringt einen individuellen Typglauben über Einheit D von 6% und über Einheit E von 38% hervor. Die Abwertung der Einheit E, die sich aus ihrer Bürgschaft für die betrügende Einheit ergibt, hat nicht nur eine Auswirkung auf den sozialen Typglauben über sie, der nunmehr 19% beträgt. Zudem vermindert sich auch der soziale Typglaube über ihren Bürgen F auf 65%. Diese Verminderung ergibt sich daraus, dass Einheit E nunmehr als weniger normativ als zuvor erscheint. Der individuelle Typglaube über die Einheit F bleibt jedoch unverändert. Dies ergibt sich zwingendermaßen daraus, dass Bürgschaftsbeziehungen nicht transitiv sind.

3. Zur Berücksichtigung der negative Empfehlung von Einheit B über Einheit C muss der Revisionsfaktor r_K berechnet werden. Hierfür kommt der soziale Typglaube über Einheit B zum Einsatz. Wir erhalten somit $r_K = 2,632$ und eine Abwertung des sozialen Typglaubens über Einheit C auf 41%. Dies entspricht auch dem individuellen Typglauben über Einheit C, da sie keine Bürgschaftsbeziehungen eingegangen ist. Hätte Einheit C Bürgen, so würde ihre Abwertung kleiner ausfallen. Dies illustriert die Wirkung von Bürgschaftsbeziehungen als soziale Absicherung.

4. Bei Auflösung der Bürgschaftsbeziehung zwischen Einheit D und E bleibt der individuelle Typglaube über die beiden unverändert. Dies ergibt sich daraus, dass im individuellen Typglauben Bürgschaftsbeziehungen nicht berücksichtigt wurde. Es bleibt also lediglich eine Revision des sozialen Typglaubens. Aufgrund des Wegfalls der Bürgschaftsbeziehung beträgt der soziale Typglaube über Einheit D nunmehr 6% und der über Einheit E 48%. Durch das Nichteingehen einer weiteren Bürgschaftsbeziehung mit Einheit D erscheint also Einheit E als weitaus normativer als zuvor.

8.4 Eingehen von Bürgschaften

Bürgschaften besitzen eine Doppelfunktion: Sie dienen als Signal und als Investition. Beide Funktionen kommen im vorangehenden Abschnitt in den Vorschriften der Glaubensbildung zur Geltung und müssen daher bei der Beantwortung der Frage, wann eine Einheit eine Bürgschaft eingeht, berücksichtigt werden.

Hiermit beschäftigen wir uns im Folgenden. Dazu entwerfen wir zunächst in Abschnitt 8.4.1 die Vorschriften für das Eingehen von Bürgschaften. Sie bestimmen das Bürgschaftsverhalten von normativen Einheiten. Zudem ist zu untersuchen, wie strategische Einheiten zum Eingehen von Bürgschaften stehen. Damit befasst sich Abschnitt 8.4.2. Die Untersuchung zeigt, dass sich normative Einheiten durch die Einführung von Bürgschaften von strategischen Einheiten glaubhaft abgrenzen können.

8.4.1 Verhalten normativer Einheiten

Die Vorschriften zum Eingehen von Bürgschaftsbeziehungen müssen die drei folgenden Fragen beantworten:

- Wann ist der Typglauben über eine Einheit hinreichend hoch, um mit ihr eine Bürgschaftsbeziehung einzugehen?
- Wie wird der Gültigkeitszeitraum der Bürgschaftsbeziehung festgesetzt?
- Wie spricht sich ein Paar von Einheiten über den Eingang von Bürgschaftsbeziehungen ab?

Bei der Beantwortung dieser Fragen ist zu berücksichtigen, dass die Vorschriften alleine das Verhalten der normativen Einheiten festlegen. Strategische Einheiten orientieren ihr Verhalten hingegen an Nutzenüberlegungen. Mit diesen beschäftigen wir uns erst im nachfolgenden Abschnitt.

Festsetzung des hinreichenden Typglaubens. In einer Bürgschaftsbeziehung bestätigen sich beide Partner gegenseitig, dass sie denselben Typ besitzen. Da die Vorschriften das Verhalten normativer Einheiten bestimmen, müssen sie daher dafür sorgen, dass Bürgschaftsbeziehungen nur mit Einheiten eingegangen werden, die hinreichend normativ erscheinen. Diese Einsicht hat sich bereits in Abschnitt 8.1.1 angedeutet. In Folgenden beschreiben wir, was unter hinreichend normativ zu verstehen ist.

Zunächst ergibt sich die Frage, ob der individuelle oder der soziale Typglaube als Kriterium für das Eingehen von Bürgschaftsbeziehungen herangezogen wird. In den sozialen Typglauben geht die Kenntnis um die Bürgen eines potentiellen Partners ein. Findet der soziale Typglaube als Kriterium Verwendung, so ist es daher möglich, dass eine Bürgschaftsbeziehung nur aufgrund

der Bürgen des Partners eingegangen wird. Dies ist insofern problematisch, als die Bürgschaftsbeziehungen des Partners zeitlich beschränkt sind. Dies erschwert die Bestimmung eines Gültigkeitszeitraums der einzugehenden Bürgschaftsbeziehung, da dabei der jeweilige Gültigkeitszeitraum seiner Bürgen zu berücksichtigen ist. Es gibt einen weiteren Grund, warum die Entscheidung über das Eingehen von Bürgschaftsbeziehung auf dem individuellen Typglauben basieren sollte: Durch das Heranziehen des individuellen Typglaubens wird sichergestellt, dass nur für solche Einheiten gebürgt wird, über die eigene Erfahrungen vorliegen. Dies ist im Sinne des dritten Entwurfsprinzips, da strategischen Einheiten somit keine andere Wahl bleibt, als sich wie normative Einheiten zu verhalten, um als Partner einer Bürgschaftsbeziehung angenommen zu werden. Es zeigt sich also, dass sich das Kriterium für das Eingehen von Bürgschaftsbeziehungen am individuellen Typglauben orientieren muss.

Welches Niveau des individuellen Typglaubens kann als hinreichend gelten? Die Antwort auf diese Frage ist eng daran verknüpft, wie stark Bürgschaftsbeziehungen in der Glaubensbildung gewichtet werden. Sind die dafür eingesetzten Modellparameter λ und ν nahe eins, so erhalten Bürgschaftsbeziehungen ein großes Gewicht. In diesem Fall muss die subjektive Wahrscheinlichkeit der Normativität des potentiellen Partners nahe 100% liegen. Je geringer die Modellparameter desto niedriger darf diese Wahrscheinlichkeit liegen. Im Folgenden bezeichnen wir das Niveau des individuellen Typglaubens, bei dem gerade noch eine Bürgschaftsbeziehung eingegangen wird, mit p_σ. Die Festsetzung dieses p_σ hängt gemäß unserer Überlegungen eng mit λ und ν zusammen. Letztendlich muss der Systementwerfer für jedes Anwendungsgebiet entscheiden, wie diese Parameter festzusetzen sind. Umfasst die Teilnahme einer Einheit am Informationssystem nur wenige Transaktionen, so ist ein relativ niedriges p_σ wünschenswert. Dadurch wird bewirkt, dass sich normative Einheiten noch während ihrer Teilnahme in Bürgschaftsbeziehungen aneinander binden können. Umgekehrt verhält es sich in Anwendungsgebieten, in denen jede Einheit an einer Vielzahl von Transaktionen teilnehmen kann. In diesem Fall ist ein hohes p_σ nahe 100% angebracht. Ein Beispiel für die Festlegung von p_σ, λ und ν findet sich in Abschnitt A.2.1 des Anhangs.

Festsetzen des Gültigkeitszeitraums. Bürgschaftsbeziehungen sind mit einem Gültigkeitszeitraum versehen, da sie von keinem der Partner widerrufen werden können. Bei der Wahl des Gültigkeitszeitraums ist also eine Abwägung zu treffen: Ist er zu kurz, müssen Bürgschaften immer wieder erneuert werden, wodurch ein gewisser Aufwand entsteht. Bei einem zu langem Zeitraum besteht für beide Einheiten hingegen die Gefahr, dass sie die Bindung zu einem späteren Zeitpunkt bereuen aber nicht auflösen können.

Wir erhalten daher die folgende Grundregel für die Festlegung des Gültigkeitszeitraums: Je mehr der individuelle Typglaube den Grenzglauben p_σ übersteigt, desto länger darf der Gültigkeitszeitraum gewählt werden. Dies liegt daran, dass in diesem Fall das Risiko eines nachträglichen Bereuens aufgrund der Normativität des Partners eher gering ist. Umgekehrt muss der Gültigkeitszeitraum für einen Typglauben nahe p_σ sehr kurz sein. Dann reicht nämlich eine negative Erfahrung mit dem Partner aus, um den Typglaube über ihn unter p_σ zu drücken und daher die Bürgschaftsbeziehung mit ihm zu bereuen.

Wie lässt sich diese Grundregel quantitativ umsetzen? Im Folgenden wird eine Quantifizierung hergeleitet, die auch in der Evaluation des Entwurfs in Kapitel 10 verwendet wird. Sie orientiert sich daran, wie lange eine Bürgschaftsbeziehung ohne Reue aufrechterhalten werden kann, wenn der Partner in den nachfolgenden Transaktionen durchweg betrügt. Als Eingaben werden zwei Abschätzungen benötigt. Einerseits ist die Zeit t_T, die zwischen zwei Transaktionen mit ihm im Mittel vergeht, einzuschätzen. Des Weiteren ist der Revisionsfaktor r_D zu bestimmen, der bei Be-

trugsverhalten in einem mittleren Transaktionskontext anzuwenden ist. Die a priori Abschätzung von t_T ist schwierig, da zum Entwurfszeitpunkt keine Informationen über Häufigkeit der Kooperation und Erreichbarkeit der Teilnehmer untereinander vorliegen. Daher bietet es sich an, den Parameter t_T während der Laufzeit anzupassen. Eine Möglichkeit hierfür ist, den durchschnittlichen Zeitraum zwischen den vergangenen Transaktionen mit dem potentiellen Bürgen zu verwenden.

Basierend auf diesen Abschätzungen berechnet sich der Gültigkeitszeitraum t_B wie folgt: Zunächst ist zu berechnen, nach wie vielen Transaktionen (ausgedrückt durch n) der individuelle Typglaube über den Partner auf p_σ fallen kann. Daraus lässt sich in der Folge direkt auf den Zeitraum t_B schließen. Die folgenden Formeln gehen davon aus, dass sich Einheit X Gedanken über den Gültigkeitszeitraum einer Bürgschaftsbeziehung mit Einheit Y macht:

$$\hat{p}_X(N_Y) \cdot (r_D)^n \stackrel{!}{=} \hat{p}_\sigma \quad \Longrightarrow \quad n = \log_{r_D} \frac{\hat{p}_\sigma}{\hat{p}_X(N_Y)}$$

$$t_B = t_T \cdot n = t_T \cdot \log_{r_D} \frac{\hat{p}_\sigma}{\hat{p}_X(N_Y)} \tag{8.9}$$

Aus der Formel lassen sich die entscheidenden Eigenschaften des Gültigkeitszeitraums ablesen:

- *Abhängigkeit vom Typglauben:* Wegen $r_D > 1$ muss $\hat{p}_\sigma > \hat{p}_X(N_Y)$ gelten, damit der Gültigkeitszeitraum positiv ist. Dies ist genau dann der Fall, wenn der individuelle Typglaube $p_X(N_Y)$ größer als p_σ ist. Damit wird die obige Grundregel eingehalten. Je normativer die Einheit Y erscheint, desto länger ist der Gültigkeitszeitraum, mit dem Einheit X bereit ist, für sie zu bürgen. Bei einem Typglauben nahe p_σ geht die Gültigkeitsdauer hingegen gegen null, da dann das Gefahr für nachträgliches Bereuen besonders groß ist.

- *Abhängigkeit von den Abschätzungen:* Je größer r_D desto kürzer ist die Gültigkeitsdauer. Dies ist sinnvoll, da der Revisionsfaktor r_D die Stärke einer Glaubensrevision im Falle von Betrug angibt. Außerdem ist die Gültigkeitsdauer proportional zum Zeitraum t_T zwischen zwei Transaktionen mit Einheit Y. Je öfters Transaktionen mit ihr eingegangen werden, desto schneller kann nämlich der Typglaube unter p_σ sinken. Auch dieser Zusammenhang ist daher sinnvoll.

Tabelle 8.3 gibt ein Beispiel für Gültigkeitszeiträume in Abhängigkeit der Höhe des individuellen Typglaubens. Dabei variieren wir den Grenzglauben p_σ zwischen 60% und 90%. Außerdem ist wie im Beispiel des Abschnitts 8.3.2 der Revisionsfaktor $r_D = 5,5$. Die jeweiligen Gültigkeitszeiträume sind normiert in der Zeit zwischen zwei Transaktionen angegeben. Die Berechnung der Gültigkeitszeiträume zeigt, dass sich die Einheiten sehr vorsichtig beim Eingehen von Bürgschaftsbeziehungen verhalten. Nur bei hoher Sicherheit über die Normativität des potentiellen Partners wird überhaupt eine Bürgschaftsbeziehung eingegangen, die länger als der Zeitraum zwischen zwei Transaktionen mit ihm dauert. Diese Vorsicht ist sinnvoll, da gemäß der Vorschriften zur Glaubensbildung das irrtümliche Bürgen von großem Nachteil ist.

Absprachen über das Eingehen von Bürgschaftsbeziehungen. Bisher sind wir die Frage über das Eingehen von Bürgschaftsbeziehungen aus der Sicht eines der beiden Partner angegangen. Dazu sind Vorschriften abgeleitet worden, unter welchen Umständen eine Einheit X mit einer Einheit Y eine Bürgschaftsbeziehung eingehen will und welchen Gültigkeitszeitraum sie hierfür wählt. Allerdings sind Bürgschaftsbeziehungen immer symmetrischer Natur. Das bedeutet, dass

Tabelle 8.3: Beispielhafte Berechnung von Gültigkeitszeiträumen einer Bürgschaftsbeziehung

$p_X(N_Y)$	60%	70%	80%	90%	95%	99%
$t_B(p_\sigma = 60\%)/t_T$	0	0,26	0,58	1,05	1,49	2,46
$t_B(p_\sigma = 70\%)/t_T$	–	0	0,32	0,79	1,23	2,20
$t_B(p_\sigma = 80\%)/t_T$	–	–	0	0,48	0,91	1,88
$t_B(p_\sigma = 90\%)/t_T$	–	–	–	0	0,44	1,41

sowohl Einheit X als auch Einheit Y zur Bürgschaft bereit sein müssen, damit die Bürgschaftsbeziehung zustande kommt. Außerdem müssen sich beide Einheiten über den Gültigkeitszeitraum verständigen. Wie dies zu geschehen hat, besprechen wir im Folgenden.

Angenommen Einheit X erkennt, dass sie eine Bürgschaftsbeziehung mit Einheit Y für die Gültigkeitsdauer $t_B(X)$ eingehen will. In diesem Fall kontaktiert sie Einheit Y und unterbreitet ihr den Vorschlag einer solchen Bürgschaftsbeziehung. Einheit Y besitzt drei Möglichkeiten, auf diese Anfrage zu antworten. Hierfür muss sie den Gültigkeitszeitraum $t_B(Y)$ ausrechnen, mit dem sie sich aus ihrer Sicht in einer Bürgschaftsbeziehung mit Einheit X binden möchte:

- $t_B(Y) \leq 0$: Ist der erwünschte Gültigkeitszeitraum negativ, so zeigt dies an, dass $p_Y(N_X) < p_\sigma$ gilt. Dies bedeutet, dass Einheit Y nicht für Einheit X bürgen möchte. In diesem Fall lehnt Einheit Y den Vorschlag zur Bürgschaftsbeziehung ab.

- $0 < t_B(Y) < t_B(X)$: Einheit Y ist zwar zum Eingehen der Bürgschaftsbeziehung bereit, allerdings mit einem kürzeren Gültigkeitszeitraum als von Einheit X vorgeschlagen. Beide Einheiten einigen sich daher auf $t_B(Y)$ als Gültigkeitszeitraum, den sie beide vertreten können. Anschließend stellen sie sich gemäß Abschnitt 8.2.1 gegenseitig Bürgschaften aus.

- $t_B(X) \leq t_B(Y)$: Der einzige Unterschied zum vorigen Fall liegt darin, dass diesmal der von Einheit X vorgeschlagene Gültigkeitszeitraum übernommen wird.

Abschließend ist noch zu klären, unter welchen Umständen eine Einheit für sich überprüft, ob sie Bürgschaftsbeziehungen vorschlagen möchte. Ein einfacher Ansatz hierfür ist eine periodische Überprüfung. Allerdings gibt es zwei Arten von Ereignissen, bei denen eine sofortige Überprüfung nahe liegt: **(1)** Wird eine eigene Bürgschaftsbeziehung ungültig, so meldet dies die Beweismittel- und Wissensverwaltung. Daraufhin ist zu prüfen, ob eine Erneuerung der Bürgschaftsbeziehung anzustreben ist. **(2)** Am Ende einer erfolgreichen Transaktion wird der eigene Transaktionspartner aufgewertet. Damit kann eine Überprüfung verbunden werden, ob aufgrund dessen eine Bürgschaftsbeziehung nunmehr eingegangen werden soll.

8.4.2 Strategisches Verhalten und seine Folgen

Die Vorschriften zum Eingehen von Bürgschaften werden nur von normativen Einheiten bedingungslos umgesetzt. Im Folgenden befassen wir uns daher mit der Frage, wie sich strategische Einheiten im Bezug auf diese Vorschriften verhalten. Anschließend stellen wir die Folgen für das Gesamtsystem heraus, die aus dem Bürgschaftsverhalten der strategischen Einheiten entstehen.

Strategisches Verhalten. Strategische Einheiten orientieren ihr Verhalten an Nutzenüberlegungen. Sie gehen also genau dann eine Bürgschaftsbeziehung ein, wenn dies in ihrem Interesse ist.

Gemäß Abschnitt 8.1.1 haben Bürgschaften eine Doppelfunktion als Signal und als Investition. Dies wirkt sich auf strategische Einheiten wie folgt aus:

- *Bürgschaft als Signal:* Je normativer der Bürge erscheint, desto glaubwürdiger ist seine Festlegung, dass der Gebürgte normativ ist, und desto stärker wird damit der Gebürgte aufgewertet. Da strategische Einheiten als normativ erscheinen wollen, ergibt sich somit, dass jede strategische Einheit Bürgschaften von normativen Einheiten anstrebt. Durch die Berechnungsvorschriften des sozialen Typglaubens erscheinen sie dadurch anderen Einheiten gegenüber als normativer.

- *Bürgschaft als Investition:* Wenn sich die Festlegung des Bürgen, dass der Gebürgte normativ ist, in Transaktionen mit dem Gebürgten bewahrheitet, wird auch der Bürge normativer als zuvor wahrgenommen. In diesem Sinne stellt das Ausstellen einer Bürgschaft eine Investition dar, die sich umso mehr auszahlt wie der Gebürgte sich normativ verhält. Eine strategische Einheit wird also für Einheiten bürgen, von denen sie in der Folge normatives Verhalten erwartet. Nur so zahlt sich die Bürgschaftsbeziehung als Investition für sie aus. Auch hieraus ergibt sich eine Bevorzugung von normativen Einheiten als Bürgen.

Wir erhalten somit, dass strategische Einheiten Bürgschaftsbeziehungen mit normativen Einheiten erstrebenswert finden. Aufgrund der Symmetrie von Bürgschaftsbeziehungen bekommt eine strategische Einheit dazu aber nur dann eine Chance, wenn eine normative Einheit sie fälschlicherweise für normativ hält. In dieser Hinsicht steht eine strategische Einheit vor zwei Alternativen:

- *Verzicht auf Bürgschaftsbeziehungen:* Durch Betrugsverhalten vergibt sich eine strategische Einheit die Chance, von anderen Einheiten als hinreichend normativ angesehen zu werden und somit als Partner in Bürgschaftsbeziehungen angenommen zu werden. Im Gegensatz zu anderen Einheiten kann sich die strategische Einheit damit nicht durch die Angabe ihrer Bürgen selbst empfehlen. Daraus ist zu folgern, dass die Betrugskosten aufgrund der Einführung von Bürgschaftsbeziehungen weiter ansteigen.

- *Aufbau von Bürgschaftsbeziehungen:* Eine strategische Einheit kann zunächst auf Betrugsverhalten verzichten, um Bürgschaftsbeziehungen aufzubauen. Erst wenn dies getan ist, erlaubt sie sich hin und wieder, Betrugsverhalten zu zeigen. Dieser Ansatz hat für die strategische Einheit zum Vorteil, dass sie sich durch Angabe ihrer Bürgen selbst empfehlen kann. Allerdings schwinden für sie die Gelegenheiten zu vorteilhaftem Betrugsverhalten. Als Ergebnis erhalten wir, dass eine strategische Einheit sich sehr viel mehr normativ als zuvor verhalten muss, um als normative Einheit zu erscheinen. Dies stärkt weiter das dritte Entwurfsprinzip. Die Auswirkung dieser Ausrichtung einer strategischen Einheit werden wir in der Evaluation des Kapitels 10 quantitativ untersuchen.

Auf den ersten Blick erscheint es für eine strategische Einheit als erstrebenswert, mit anderen strategischen Einheiten eine Bürgschaftsbeziehung einzugehen. Dadurch ließe sich trotz eigenen Betrugsverhaltens eine Reihe von Einheiten als Bürgen in einer Selbstempfehlung angeben. Das Problem liegt aber darin, dass weniger die Zahl als die Qualität der Bürgen Ausschlag gibt. Erscheinen die eigenen Bürgen als strategisch, so erzielt eine Selbstempfehlung eine negative Wirkung. Hinzu kommt, dass es gefährlich ist, für andere strategische Einheiten zu bürgen. Diese sind sich nämlich darüber bewusst, dass sich eine Bürgschaft gemäß Abschnitt 8.3.2 wie eine Versicherung auswirkt. Die gebürgten strategischen Einheiten nehmen daher den so genannten

Moral Hazard[7] wahr, dass sie die negativen Folgen ihres Betrugsverhalten zum Teil auf ihre Bürgen abwälzen können. Als Folge von alledem ist eine Bürgschaftsbeziehung mit strategisch erscheinenden Einheiten nicht wünschenswert. Schlussendlich erhalten wir somit, dass strategische Einheiten nur Bürgschaftsbeziehungen mit normativen Einheiten erstrebenswert finden.

Folgen für das Gesamtsystem. Die Untersuchung des vorigen Paragraphs hat gezeigt, dass auch strategische Einheiten nur auf Bürgschaftsbeziehungen mit normativen Einheiten aus sind. Dies verschärft die soziale Kontrolle zwischen den Einheiten: Einerseits steigen die Betrugskosten, da durch Betrugsverhalten die Möglichkeit zu Bürgschaftsbeziehungen unterminiert wird. Andererseits ist ein viel höheres Maß an kooperativem Verhalten vonnöten, damit eine strategische Einheit ihr Ziel erreicht, als normativ angesehen zu werden. Auch dies liegt an der Einführung von Bürgschaftsbeziehungen. Da normative Einheiten als Bürgen umworben sind, müssen strategische Einheiten sich quasi so wie die normativen Einheiten verhalten, um zu Bürgschaftsbeziehungen zu kommen. Dies führt für strategische Einheiten zu ähnlichen Opportunitätskosten, wie sie normativen Einheiten tragen. Aus der Sicht der Signalisierungstheorie erreichen wir durch die Einführung von Bürgschaftsbeziehungen somit eine Erhöhung der Kosten für die Signalstörung. Als Ergebnis erhalten wir eine erhöhte Unterscheidbarkeit zwischen normativen und strategischen Einheiten. Die Erweiterung des Entwurfs in diesem Kapitel führt also dazu, dass sich normative Einheiten durch die Zahl und die Qualität ihrer Bürgen von strategischen Einheiten absetzen können.

Wie ist die Möglichkeit der strategischen Einheiten zu bewerten, durch sehr normatives Verhalten dennoch zu Bürgschaften zu kommen? Der weitgehende Verzicht auf Betrugsverhalten bringt mit sich, dass strategische Einheiten kaum einen Vorteil gegenüber normativen Einheiten genießen. In der Terminologie des Systementwurfs aus Kapitel 5 werden damit die Normativitätskosten sehr gering. Dies wirkt sich auf die Typwahl der menschlichen Prinzipale aus: Durch die erhebliche Absenkung der Normativitätskosten sind weitaus weniger Benutzer dazu bereit, die Risiken und Nachteile zu tragen, die eine Verwendung einer manipulierten Version der Systemsoftware mit sich bringt. In letzter Instanz führt die Einführung von Bürgschaftsbeziehungen daher zu einer Erhöhung des Anteils derjenigen Benutzer, die sich zur Verwendung der originalen Systemsoftware entscheiden. Dies trägt zur Erreichung unseres Zieles bei, die Existenzfähigkeit des Informationssystems zu gewährleisten.

8.5 Zusammenfassung

Normative Einheiten benötigen eine Möglichkeit zum positiven Empfehlen, um sich gezielt von strategischen Einheiten absetzen zu können. In diesem Kapitel haben wir daher den Entwurf um soziale Beweismittel erweitert, durch die sowohl aussagekräftige als auch glaubwürdige Selbstempfehlungen ermöglicht werden. Hierfür sind wir zunächst auf das Konzept der *sozialen Bindungen* eingegangen, das den sozialen Beweismitteln zugrunde liegt. Soziale Bindungen machen das soziale Gefüge explizit und erlauben dadurch, dass aus der Stellung einer Einheit im sozialen Gefüge Aufschlüsse über ihren Typ gemacht werden können. Wir haben den Entwurfsraum für soziale Bindungen untersucht und die Typ-Bürgschaftsbeziehung als die Art der sozialen Bindung identifiziert, die einem Informationssystem wie dem des Campus-Szenarios am angemessensten ist.

[7] Mit Moral Hazard bezeichnet man die Neigung einer autonomen Einheit, ihr Verhalten nach Abschluss einer Versicherung anzupassen [BS89]. Dies wird in der menschlichen Gesellschaft zum Beispiel beobachtet, wenn Autofahrer nach Abschluss einer Versicherung riskanter fahren, da sie von den negativen Folgen eines Unfalls durch den Versicherer zum Teil entlastet werden.

Zwei Einheiten gehen eine solche Bindung ein, indem sie sich gegenseitig eine Bürgschaft in Form eines Beweismittels ausstellen. Bei einer Bürgschaft handelt es sich um ein soziales Beweismittel, da darin die soziale Bindung festgehalten wird.

Der Einsatz der sozialen Beweismittel erfordert eine Erweiterung des Kreislaufs der verteilten Vertrauensbildung. Wir haben eine Übersicht über diese Erweiterung gegeben und den Lebenszyklus eines sozialen Beweismittels beschrieben. Bürgschaften werden im Rahmen des Sechs-Wege Transaktionsprotokolls *ausgestellt*. Dadurch wird der ohnehin sehr geringe Betrugsvorteil beim Eingehen einer Bürgschaftsbeziehung weiter vermindert. Das *Empfehlungssystem* wurde um die Möglichkeit erweitert, sich durch die Angabe der eigenen Bürgen selbst zu empfehlen. Die dazu vorgewiesenen Bürgschaften führen zur Aussagekraft und Glaubwürdigkeit von Selbstempfehlungen. Die Aspekte, die für die Einbeziehung von Bürgschaften in die *Beweismittel- und Wissensverwaltung* zu berücksichtigen sind, wurden ebenfalls besprochen.

Die *Glaubensbildung* wurde um die Ebene des sozialen Typglaubens erweitert, in die die Kenntnis von Bürgschaftsbeziehungen Anderer einfließt. Im Gegensatz zum individuellen Typglauben aus den vorangegangen Kapiteln speichert eine Einheit ihren sozialen Typglauben nicht explizit, sondern leitet ihn nach Bedarf aus dem individuellen Typglauben und den bekannten Bürgschaftsbeziehungen ab. Dafür wurden Berechnungsvorschriften vorgestellt, in denen die Gewichtung der Bürgschaften parametrisierbar ist. Anschließend haben wir ein Modell entwickelt, mit dem der Zusammenhang zwischen individuellem und sozialem Typglauben bei der Durchführung von Glaubensrevisionen gewahrt wird. Auch dieses Modell ist parametrisierbar, so dass der Glauben über eine Einheit und über ihre Bürgen unterschiedlich stark revidiert werden kann. Die Funktionsweise der Vorschriften der Glaubensbildung und -revision wurde in einem ausführlichen Beispiel dargestellt.

Abschließend haben wir Vorschriften dafür abgeleitet, wann Einheiten zum *Eingehen* von Bürgschaftsbeziehungen bereit sind. Die Analyse zeigt, dass nicht nur normative sondern auch strategische Einheiten bevorzugt mit normativen Einheiten Bürgschaftsbeziehungen eingehen. Dies liegt daran, dass das Ausstellen einer Bürgschaft eine Investition darstellt, die sich nur bei kooperativem Verhalten des Gebürgten auszahlt. Schlussendlich haben wir somit erreicht, dass sich normative Einheiten durch die Zahl und der Qualität der eigenen Bürgen glaubhaft von strategischen Einheiten abgrenzen können.

Teil III

Evaluation

Kapitel 9

君子喻于義 小人喻于利

"Der Edle setzt sich mit der Frage auseinander, was richtig ist, während sich der Niederträchtige damit befasst, was sich für ihn am meisten auszahlt."
(Gespräche und Aussprüche des Konfuzius, 4.16)

Methodik der Evaluation

Diese Arbeit geht der Frage nach, ob die Vision von selbstorganisierenden Informationssystemen wie dasjenige des Campus-Szenarios realisiert werden kann. Die Existenzfähigkeit dieser Informationssysteme wird von der Autonomie der Teilnehmer und der daraus folgenden Möglichkeit zur Manipulation bedroht. Um die Vorteile von Betrugsverhalten einzuschränken, wurden hierfür in Teil II der Arbeit Verfahren der verteilten Vertrauensbildung entworfen, die für soziale Kontrolle der einzelnen Einheiten untereinander sorgen. Die Analysen der Kapitel 7 und 8 haben zwar gezeigt, dass der Einsatz von Beweismitteln zur Verschärfung dieser sozialen Kontrollen führt. Ein abschließendes Urteil zur Existenzfähigkeit des Informationssystems kann jedoch nur im Zuge einer umfassenden Evaluation erfolgen. Sie ist für eine quantitative Bewertung des Entwurfes zuständig. Mit dieser Aufgabe befassen wir uns in diesem Teil der Arbeit.

Die Grundlage für die Evaluation des Entwurfs wird in diesem Kapitel gelegt. Es entwickelt eine Methodik der simulativen Evaluation, deren Ergebnisse Aussagekraft über das Zutreffen unserer These über die Existenzfähigkeit des Informationssystems besitzen. Abschnitt 9.1 führt in die Grundkonzepte dieser Methodik ein. Die simulative Evaluation verlangt nach Computerunterstützung, die durch das Entwickeln entsprechender Simulationswerkzeuge erbracht wird. Abschnitt 9.2 bespricht das Simulative Kooperationsturnier als das Werkzeug, das letztendlich zur Simulation des Gesamtsystems eingesetzt wird. Es wird durch das Interaktive Kooperationsturnier aus Abschnitt 9.3 ergänzt. Es bildet die Grundlage dafür, die Richtungen der Manipulation zu antizipieren.

9.1 Einführung

Die Methodik zieht die Simulation als das wesentliche Mittel der Evaluation heran. Abschnitt 9.1.1 bespricht das Konzept, das dieser simulativen Evaluation zugrunde liegt. Die Evaluation orientiert sich an unserem Ziel, die These von der Existenzfähigkeit des Informationssystems zu verifizieren. Die sich dadurch ergebende Ausrichtung der Evaluation wird in Abschnitt 9.1.2 besprochen.

Diese Vorbetrachtungen führen in Abschnitt 9.1.3 zur Festlegung eines angemessenen Evaluationsprozesses. In ihm wird deutlich, welche Formen der computergestützten Simulation für die Evaluation des Entwurfes erforderlich sind.

9.1.1 Konzept der simulativen Evaluation

Im Folgenden wird das Konzept vorgestellt, das der simulativen Evaluation zugrunde liegt. Dazu vergleichen wir zunächst ihre Eigenschaften mit denen der analytischen Evaluation und zeigen, dass die Simulation die angemessenere Methode zur Evaluation darstellt. Anschließend befassen wir uns mit den Elementen, die in die simulative Evaluation eingehen.

Simulative oder analytische Evaluation. Abschnitt 2.4.2 nennt zwei Alternativen, mit denen die Robustheit einer verteilten Vertrauensbildung evaluiert werden kann:

- In der *simulativen Evaluation* werden die Abläufe des Gesamtsystem mitsamt der Verhaltensweisen der Einheiten und den Rahmenbedingungen, die für das System gelten, auf einem Computer nachgebildet. Da sich das Ergebnis einer Simulation somit stets nur auf ganz bestimmte Rahmenbedingungen bezieht, sind eine Reihe von Durchläufen unter leicht veränderten Rahmenbedingungen erforderlich, um allgemeine Aussagen über das Gesamtsystem treffen zu können. Weiterhin erfordert eine realitätsnahe Simulation, dass die Möglichkeiten zu Fehlverhalten antizipiert werden und die Nachbildung der Rahmenbedingungen des Informationssystems wirklichkeitsgetreu ist.

- Die *analytische Evaluation* orientiert sich hingegen an den analytischen Verfahren der Spieltheorie. Zu untersuchende Entscheidungssituationen werden hierfür als Spiel modelliert. Ausgehend von dessen Modelleigenschaften kommt die Analyse auf deduktive Weise zu Aussagen über das Gesamtsystem. Soweit das Modell die entscheidenden Aspekte der Realität richtig erfasst, haben somit die Ergebnisse der analytischen Evaluation eine größere Aussagekraft als die der simulativen Evaluation. Das Problem liegt jedoch darin, alle Einflussfaktoren im Vorhinein zu erkennen und auf angemessene Weise in das Modell aufzunehmen.

Die Gegenüberstellung der beiden Alternativen der Evaluation zeigt die Überlegenheit der simulativen Evaluation. Wie Axelrod in seinem Plädoyer für die simulative Evaluation darlegt [Axe97], führt die Simulation in komplexen Umgebungen wie der des Campus-Szenarios zu aussagekräftigeren Ergebnissen als die Analyse. Dies liegt daran, dass beim Aufstellen des Modells, das der Analyse zugrunde liegt, notwendigerweise starke Vereinfachungen der Rahmenbedingungen vorgenommen werden müssen. Außerdem ist das Zusammenspiel des Verhaltens der einzelnen Einheiten untereinander derart komplex und unter Berücksichtigung der Umgebung indeterministisch, dass die Beschränkung auf einzelne Entscheidungssituationen die Aussagekraft der analytischen Evaluation weiter einschränkt. Somit ergibt sich, dass die Simulation die angemessenere Methode der Evaluation darstellt.

Elemente der simulativen Evaluation. Auf der konzeptionellen Ebene lassen sich verschiedene Elemente unterscheiden, die in die simulative Evaluation eingehen. Im Folgenden gehen wir nacheinander auf diese ein.

Wie bereits angesprochen haben die *Rahmenbedingungen*, die für das Gesamtsystem gelten, einen entscheidenden Einfluss auf dessen Eigenschaften. Das Systemmodell aus Abschnitt 1.2.2

gibt einen Anhaltspunkt darüber, was im Einzelnen unter den Rahmenbedingungen zu verstehen ist. Sie werden wie das Systemmodell durch vier Teile bestimmt: **(1)** *Einheiten:* Aus wie vielen Einheiten besteht das Informationssystem? Wie hoch ist der Anteil normativer Einheiten? Welche manipulierte Version der Systemsoftware wird jeweils von den strategischen Einheiten verwendet? **(2)** *Kommunikation:* Welche Einheiten können zu welchen Zeitpunkten untereinander kommunizieren? Wie hoch ist dabei die Wahrscheinlichkeit von Kommunikationsabbrüchen? **(3)** *Kooperation:* Wie werden die Vorteile der Kooperation im Vergleich zu ihren Nachteilen eingeschätzt? **(4)** *Transaktionsgelegenheiten:* Wann bieten sich Gelegenheiten zu beidseitig vorteilhaften Transaktionen? Gibt es dabei unterschiedliche Möglichkeiten bei der Wahl des Transaktionspartners?

Aus der Sicht des Systementwerfers lassen sich somit die Rahmenbedingungen als die Gegebenheiten charakterisieren, die von außen vorgegeben sind. Im Gegensatz dazu kann der Systementwerfer gemäß Abschnitt 5.1 lediglich über die Eigenschaften der originalen Systemsoftware und damit das Verhalten der normativen Einheiten bestimmen. Dieser Gestaltungsraum wird in der Theorie des Systementwurfs *Entwurfsraum* genannt [Gri03]. Bei der Durchführung des Entwurfs in den Kapiteln 6 bis 8 haben die Normen und Vorschriften für die Festlegung des Verhaltens der normativen Einheiten gesorgt. In dieser Hinsicht stellt der Entwurf dieser Arbeit einen *Entwurfspunkt* dar. Dadurch, dass einige Vorschriften der Glaubensbildung parametrisierbar sind, ist dieser Entwurfspunkt nicht eindeutig bestimmt. Wie bereits im Entwurf gezeigt, hängt die Quantifizierung dieser Parameter von der Anwendungsdomäne ab. Eine Bestimmung der Parameter im Sinne des Campus-Szenarios ermöglicht daher die Festlegung des Entwurfspunktes.

Im Zuge der simulativen Evaluation finden pro Simulationslauf jeweils genau ein Entwurfspunkt und eine Festlegung der Rahmenbedingungen (die so genannte *Benchmark* [Gri03]) Eingang. Damit die Simulation von Nutzen ist, müssen während der Simulation einige Eigenschaften des Systems festgehalten werden. Dies führt uns zum dritten Element der simulativen Evaluation, den *Metriken*. Eine Metrik gibt eine quantitative Aussage über die Eigenschaften des Systems. Daher bilden Metriken die Grundlage für die Interpretation des Simulationsausgangs. Die Art der Metriken und der Vorgang der Interpretation werden dabei alleine von der Ausrichtung der simulativen Evaluation, also ihrem Ziel, bestimmt.

Der Zusammenhang zwischen den Elementen der simulativen Evaluation wird in Abbildung 9.1 zusammengefasst.

9.1.2 Ausrichtung der Evaluation

Das Ziel des Entwurfs der verteilten Vertrauensbildung ist es, die Existenzfähigkeit des Informationssystems zu gewährleisten und damit die These dieser Arbeit zu untermauern. Eine Evaluation des Entwurfs muss also daran ausgerichtet sein, die Erreichung dieses Zieles zu überprüfen.

Im Folgenden untersuchen wir, was im Einzelnen die Folgen dieser Ausrichtung sind. Dabei entwickeln wir zunächst ein Kriterium, mit Hilfe dessen Aussagen über die Existenzfähigkeit des Informationssystems getroffen werden können. Anschließend wird eine Maßzahl identifiziert, die die Güte des Entwurfs und den Grad der Zielerreichung festhält. Dies ermöglicht eine Beschreibung davon, wie eine Simulation durchgeführt wird und ihre Ergebnisse zu interpretieren sind.

Kriterium für die Existenzfähigkeit. Die Existenzfähigkeit des Informationssystems wird von zwei Seiten bedroht. Einerseits können sich Benutzer, die bisher die originale Systemsoftware verwendet haben, zur Installation einer manipulierten Version davon entscheiden. Andererseits können enttäuschte Benutzer aus dem System austreten. Laut der Analyse des Abschnitts 5.3.2

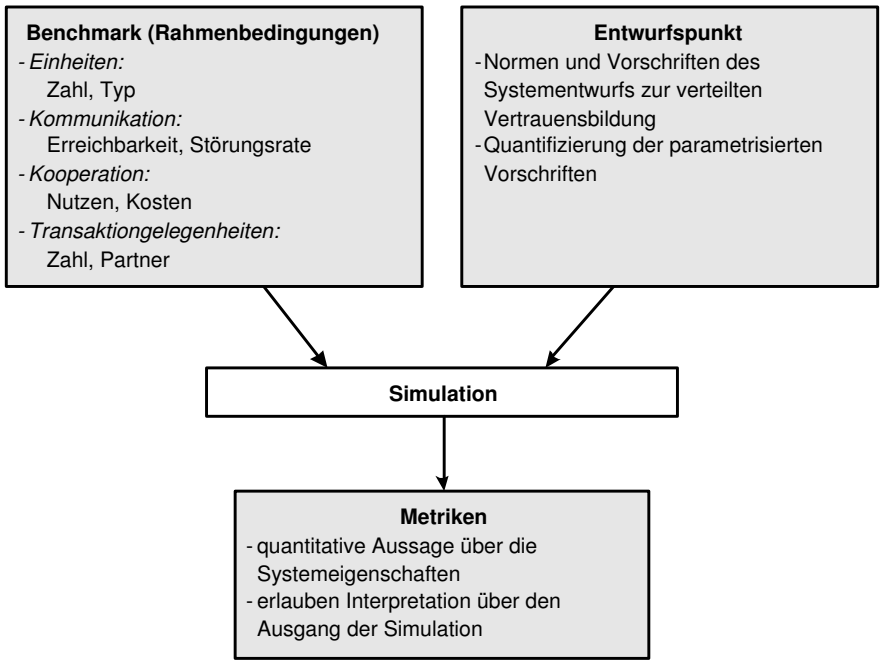

Abbildung 9.1: Die Elemente der simulativen Evaluation

kommt es aus diesen beiden Gründen zur Degeneration des Informationssystems und in der Folge zum Aufhören seiner Existenz.

Wenn wir ein Kriterium für die Existenzfähigkeit des Informationssystems entwickeln wollen, müssen wir uns also daran orientieren, dass es nicht zu einer solchen Degeneration kommen darf. Eine hinreichende Bedingung hierfür ist die Stabilitätsbedingung aus Abschnitt 5.3.2, die für jeden Attraktor des Gesamtsystems gelten muss. Sie ist genau dann erfüllt, wenn sich jeder Benutzer nach Beobachtung der Systemeigenschaften in der Wahl des Typs seiner Einheit bestätigt fühlt. Diese Bedingung ist insofern hinreichend für die Existenzfähigkeit des Informationssystems, als ein stabiles System nicht degenerieren kann.

Das Kriterium für die Existenzfähigkeit des Informationssystems stellt eine abgeschwächte Form der Stabilitätsbedingung dar. Wir müssen in diesem Kriterium lediglich fordern, dass es nicht zur Degeneration kommt. Der Unterschied zur Stabilitätsbedingung ergibt sich daraus, dass nur für die Prinzipale normativer Einheiten gefordert wird, dass sie ihre Typwahl nicht ändern. Dies ist ausreichend für die Sicherstellung davon, dass keiner der zwei Gründe der Degeneration zutrifft. In Anlehnung an die Formulierung der Stabilitätsbedingung erhalten wir somit das folgende Kriterium:

- *Teilnahme:* Die Prinzipale der normativen Einheiten wollen weiterhin am System teilnehmen. Dieses Kriterium lässt sich anhand des Individualnutzens der jeweiligen Benutzer überprüfen.

- *Normativität:* Die Prinzipale der normativen Einheiten bleiben bei der originalen Version der Systemsoftware. Für sie gilt daher, dass ihre jeweiligen Manipulationskosten die Normativitätskosten übersteigen.

Maßzahl für die Güte des Entwurfs. Aus der Sicht des Systementwerfers ist es zwar erforderlich, die Existenzfähigkeit des Informationssystems herzustellen. Darüber hinaus ist jedoch sein Ziel, dass so viele Benutzer wie möglich mit der originalen Systemsoftware am Informationssystem teilnehmen. Nur so können die Systemeigenschaften vom Entwerfer bestimmt werden. Das Kriterium für die Existenzfähigkeit macht aber über den Anteil der normativen Einheiten am System keine Aussage. Daher bedarf es einer Maßzahl, die die Güte des Entwurfs festhält.

Bei der Besprechung der Ausrichtung des Entwurfes haben wir in Abschnitt 5.4.1 ein Kriterium abgeleitet, das die Güte des Entwurfs angibt. Der Systementwurf ist umso erfolgreicher, je höher der Anteil der Benutzer ist, die am System mit der originalen Systemsoftware teilnehmen. Voraussetzung dafür ist, dass zusätzlich das Kriterium für die Existenzfähigkeit des Systems erfüllt ist. Dann verbleiben nämlich diese Benutzer mit der originalen Systemsoftware im System. Der Anteil normativer Einheiten am System ist dann die Maßzahl, die die Güte des Entwurfs anzeigt.

Vorgang und Interpretation der Simulation. Wie sind diese Überlegungen in der Durchführung der Simulation zu berücksichtigen? Zunächst ist festzuhalten, dass die Größen, die Eingang in das Kriterium finden, unterschiedlichen Elementen der simulativen Evaluation angehören. Auf der einen Seite werden der Anteil normativer Einheiten und die Verteilung der Manipulationskosten im Rahmen der Benchmark vorgegeben, da dies die Voraussetzung für einen Simulationslauf ist. Auf der anderen Seite werden die Normativitätskosten und der Individualnutzen der Benutzer gemessen und in entsprechenden Metriken festgehalten. Anhand dieser Metriken lässt sich direkt ablesen, ob das Kriterium der Existenzfähigkeit eingehalten worden ist.

Ein vollständiges Bild der Eigenschaften des Informationssystems erreichen wir, wenn wir die Simulation unter variierten Rahmenbedingungen mehrfach durchführen. Dabei ist die Variation des Anteils normativer Einheiten von besonderem Interesse. Hierbei werden wir von der Frage geleitet, bei welchem Anteil normativer Einheiten das Informationssystem existenzfähig ist. Ein möglicher Verlauf der Messungen wird in Abbildung 9.2 dargestellt. Bei ansonsten fest vorgegebenen Rahmenbedingungen wird der Anteil normativer Einheiten im Bereich zwischen 15% und 95% variiert. Ist das Kriterium der Existenzfähigkeit erfüllt, so markiert die Abbildung den entsprechenden Punkt beispielhaft mit einer ausgefüllten (grünen) Raute. Andernfalls bleibt die Raute unausgefüllt (und rot). In der Abbildung bilden die Punkte mit ausgefüllten (grünen) Rauten einen zusammenhängenden Bereich zwischen 15% und 90%, der entsprechend eingefärbt ist. Dies ist der Bereich, in dem das Informationssystem existenzfähig ist. Wir sprechen in dieser Hinsicht auch von dem *Gültigkeitsbereich*, da wir in diesem Bereich unsere These von der Existenzfähigkeit des Informationssystems validieren können. Für einen Anteil normativer Einheiten über 90% ist das Kriterium hingegen nicht erfüllt. Das in der Abbildung dargestellte Ergebnis ist typisch für die Evaluation. Die Gründe hierfür sind bereits in Abschnitt 5.4.1 angesprochen worden. Das beispielhafte Ergebnis rührt daher, dass wir bei der Verteilung der Manipulationskosten davon ausgehen müssen, dass es einige wenige Benutzer gibt, denen die Manipulation der originalen Systemsoftware kaum oder keine Kosten verursacht. Diese inhärent manipulationsfreudigen Benutzer von der Verwendung der originalen Systemsoftware zu überzeugen, ist daher äußerst schwierig oder unmöglich.

Für die Beurteilung der Güte des Entwurfes bedeutet das Ergebnis dieser Abbildung das Folgende: Bei einem hohen Anteil normativer Einheiten (bis zu 90%) ist das Informationssystem existenzfähig. Die Degeneration des Informationssystems tritt nur anfangs unter der Einwirkung einiger weniger inhärent manipulationsfreudiger Benutzer ein. Bei einem Anteil normativer Einheiten um 90% findet sich jedoch kein Benutzer mehr, der von der originalen Systemsoftware

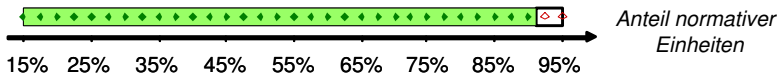

Abbildung 9.2: Beispielhafte Simulation mit variiertem Anteil normativer Einheiten

auf eine manipulierte Version von ihr umsteigen möchte[1]. Dies bedeutet unter anderem, dass die Normativitätskosten gering sind. Ansonsten wäre es zu einer weiteren Verbreitung manipulierter Versionen gekommen. Niedrige Normativitätskosten bedeuten, dass die inhärent manipulationsfreudigen Benutzer gegenüber diejenigen Benutzer, die die originale Systemsoftware benutzen, kaum einen Vorteil haben. Insofern ist es auch unter dem Gesichtspunkt der Fairness tragbar, dass nicht alle Einheiten normativ sind. Somit setzt sich die originale Systemsoftware im Gesamtsystem durch.

Da das Ergebnis der Simulation auch maßgeblich von anderen Bestandteilen der Rahmenbedingungen abhängt, reicht eine Variation des Anteils normativer Einheiten in der Durchführung der Simulation nicht aus. Darüber hinaus müssen weitere Faktoren wie zum Beispiel die Wahrscheinlichkeit für Kommunikationsabbrüche oder die Häufigkeit der Gelegenheiten zum Eingehen einer Transaktion variiert werden.

Wir fassen zusammen, dass wir zwar ein einfaches Kriterium besitzen, um die Existenzfähigkeit des Informationssystems und die Güte des Entwurfes zu beurteilen. Jedoch sind eine ganze Reihe von Simulationsläufen vonnöten, um die Abhängigkeit der Erfüllung des Kriteriums von den Eigenheiten der Rahmenbedingungen abschätzen zu können.

9.1.3 Evaluationsprozess

In diesem Abschnitt wird ein Prozess vorgestellt, mit Hilfe dessen die Evaluation des eigenen Ansatzes erfolgt. Hierfür geben wir im Folgenden eine Übersicht des gesamten Prozesses und der Werkzeuge, die dabei unabdingbar sind.

Bei der simulativen Evaluation des eigenen Ansatzes ergibt sich die Frage, wie das Verhalten strategischer Einheiten zu modellieren ist. Diese Frage lässt sich darauf zurückführen, welche manipulierten Versionen der Systemsoftware in einem realen Informationssystem erstellt werden und für welche dieser Versionen sich ein menschlicher Prinzipal, der zur Manipulation bereit ist, entscheidet. Eine Beantwortung dieser Frage ist keineswegs trivial, da hierzu die Angriffspunkte gegen das normative Verhalten, wie es im eigenen Ansatz vorgesehen ist, systematisch aufgedeckt werden müssen. Die Strategie, mit der eine manipulierte Version der Systemsoftware diese Angriffspunkte ausnutzt und damit den Systementwurf durchkreuzt, nennen wir im Folgenden *Gegenstrategie*. Es ist durchaus denkbar, dass mehrere Gegenstrategien angewendet und somit verschiedene manipulierte Versionen der Systemsoftware erstellt werden. Die erste Aufgabe der Evaluation des eigenen Ansatzes besteht also in der *Antizipation der Manipulation*. Ergebnis dieser Antizipation ist eine Abbildung zwischen den Rahmenbedingungen und den darin zu erwartenden Gegenstrategien.

Im Evaluationsprozess gliedert sich die Antizipation der Manipulation in zwei Schritte: **(1)** Zuerst ist der Raum der Gegenstrategien auf systematische Weise zu durchsuchen und viel versprechenden Gegenstrategien sind zu *finden*. Das Ergebnis dieses Schrittes ist die Identifikation und Formulierung der Gegenstrategien. **(2)** Anschließend ist für unterschiedliche Rahmenbedin-

[1]Diesen dynamischen Verlauf der Populationsstruktur haben wir bereits in Abschnitt 5.3.2 gekennzeichnet. Gemäß Abbildung 5.5 würde der Fixpunkt von $F(N(b))$ bei $b^* = 10\%$ liegen.

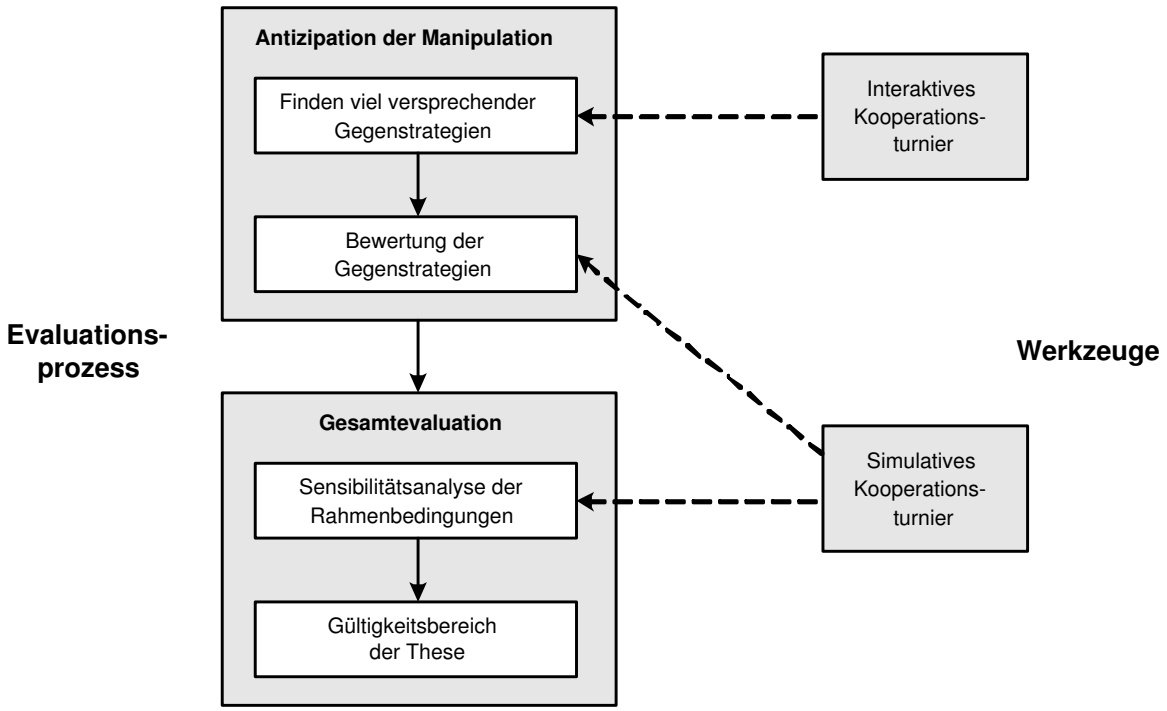

Abbildung 9.3: Der Evaluationsprozesses

gungen zu ermitteln, welche der Gegenstrategien am besten abschneiden und daher zu antizipieren sind. Aus dieser *Bewertung* der Gegenstrategien wird die Abbildung zwischen den Rahmenbedingungen und den darin zu erwartenden Gegenstrategien abgeleitet. Damit sind wir in der Lage, bei der Gesamtevaluation die jeweils beste Gegenstrategie in Abhängigkeit der zu simulierenden Rahmenbedingungen für die strategischen Einheiten auszuwählen.

Erst nach der Antizipation der Manipulation kann es zur Evaluation des Gesamtsystems kommen. Wie bereits im vorhergehenden Abschnitt gezeigt, müssen die Rahmenbedingungen für die Evaluation des Gesamtsystems variiert werden. Nur so erhalten wir Aussagen über den Gültigkeitsbereich unserer These von der Existenzfähigkeit des Informationssystems.

Abbildung 9.3 zeigt den gesamten Evaluationsprozess, der sich aus diesen Überlegungen ergibt. Er erfordert die Verwendung von zwei Werkzeugen:

- *Interaktives Kooperationsturnier:* Dieses Werkzeug ist zur systematischen Durchsuchung des Raums der Gegenstrategien zuständig. Die Namensgebung dieses Werkzeuges erklärt sich daraus, dass menschliche Versuchspersonen zu diesem Zweck herangezogen werden. Abschnitt 9.3 befasst sich mit diesem Werkzeug.

- *Simulatives Kooperationsturnier:* Die eigentliche simulative Evaluation erfolgt mit diesem zweiten Werkzeug. Es ermöglicht die automatisierte Durchführung der Simulation und ihre Auswertung. Um die Abhängigkeit der Simulationsergebnisse von den Rahmenbedingungen bewerten zu können, setzt dieses Werkzeug eine Sensibilitätsanalyse ein. Das simulative Kooperationsturnier eignet sich nicht nur für die Evaluation des Gesamtsystems. Darüber hinaus wird es auch zur Bewertung der Gegenstrategien herangezogen. Der nachfolgende Abschnitt 9.2 befasst sich mit diesem Werkzeug.

9.2 Simulatives Kooperationsturnier

Der Evaluationsprozess sieht das Simulative Kooperationsturnier als ein Werkzeug vor, das für die Evaluation des Gesamtsystems zuständig ist. Die hierfür nötige Simulationsumgebung wird in Abschnitt 9.2.1 besprochen. Anschließend befassen wir uns in Abschnitt 9.2.2 mit der Einbindung einer Sensibilitätsanalyse, die den automatisierten Test des Gesamtsystems unter verschiedenen Rahmenbedingungen erlaubt.

9.2.1 Simulationsumgebung

Da sich die simulative Evaluation am Campus-Szenario orientiert, sind bei der Entwicklung der Simulationsumgebung die Eigenheiten dieses Szenarios zu berücksichtigen. In diesem Abschnitt besprechen wir zunächst das Rahmenwerk der Simulation, auf das das Simulative Kooperationsturnier aufsetzt. Anschließend befassen wir uns mit der Einbindung der Rahmenbedingungen des Informationssystems, soweit diese im Rahmenwerk unberücksichtigt geblieben sind.

Rahmenwerk der Simulation. Ausgangspunkt der computergestützten Simulation sind generische Simulationswerkzeuge, die unabhängig von der Anwendungsdomäne für die Grundlagen der Simulation sorgen. Zu diesen Grundlagen zählen die erweiterbare Modellierung eines verteilten Systems, die Bereitstellung einer Simulationsoberfläche und -steuerung und ein System zur Berechnung von Metriken. In dieser Hinsicht handelt es sich um ein Rahmenwerk der Simulation, das für die eigentlichen Bedürfnisse der simulativen Evaluation zu spezialisieren ist.

DIANEmu [Kle03] ist ein solches Rahmenwerk der Simulation. Der entscheidende Vorteil dieses Rahmenwerks ist, dass es Aspekte des Campus-Szenarios erfasst. Abbildung 9.4 zeigt eine Momentaufnahme der Simulationsoberfläche. Darin ist ein Teil des Campus der Universität Karlsruhe dargestellt. Die Positionen der einzelnen Informationsgeräte auf dem Campus sind als ausgefüllte Kreise dargestellt. In DIANEmu ist bereits ein Modell integriert, das die Bewegung der Studenten auf dem Campus-Gelände erfasst. Dieses Bewegungsmodell ist anhand von Studentenbefragungen entwickelt und in [BKOKR04, KRKB04] vorgestellt worden.

Im Rahmen dieser Arbeit wurde DIANEmu zum Simulativen Kooperationsturnier ausgebaut. Hierfür wurden die Elemente der simulativen Evaluation auf die Ausrichtung dieser Arbeit abgestimmt. Konkret bedeutet dies, dass der Entwurf des Teils II dieser Arbeit und die Metriken, anhand derer gemäß Abschnitt 9.1.2 die Simulationsergebnisse interpretiert werden, in DIANEmu implementiert worden sind. Zusätzlich sind für das Simulative Kooperationsturnier auch die Rahmenbedingungen des Informationssystems, die über das Bewegungsmodell hinausgehen, berücksichtigt und modelliert worden. Mit dieser Modellierung der Rahmenbedingungen befasst sich der nachfolgende Paragraph. Als Ergebnis erhalten wir das Simulative Kooperationsturnier als ein Simulationswerkzeug, das den eigenen Ansatz in der Umgebung des Campus-Szenarios realitätsnah simulieren kann und die Simulationsergebnisse in Form der benötigten Metriken zur Verfügung stellt.

Modellierung der Rahmenbedingungen des Informationssystems. Im Folgenden wenden wir uns der Modellierung der Rahmenbedingungen zu. Hierfür stellen wir je ein Modell für die Bestandteile der Rahmenbedingungen gemäß Abschnitt 9.1.1 auf. Zunächst befassen wir uns mit dem Modell der Kooperation, das auch die Transaktionen umfasst. Anschließend erörtern wir, wie ein Modell für die Kommunikation und für die Einheiten aussieht.

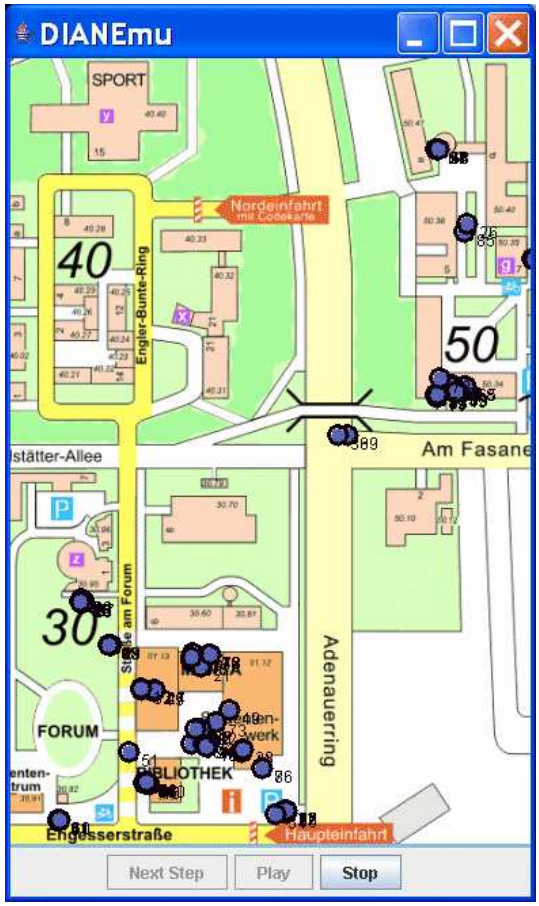

Abbildung 9.4: Darstellung des Campus-Szenarios im Simulationsrahmenwerk DIANEmu

Modell der Kooperation. Um ein Modell zur Kooperation zu entwickeln, gehen wir auf den typischen Ablauf des Campus-Szenarios aus Abschnitt 1.2.1 zurück: In dem Moment, in dem Anna ihren Vorlesungsmitschrieb fertiggestellt hat, ist ihr Gerät in Besitz einer Information, die auch von den menschlichen Prinzipalen anderer Geräte benötigt wird. Anna selbst interessiert sich für den Mensaplan des Tages. Durch die Fertigstellung des Vorlesungsmitschriebs gerät ihre Einheit also in eine Situation, in der sie eine *Gelegenheit* zur Transaktion erhält. In der Szenariobeschreibung gibt es mit Claude und David zwei Studenten, die den Mensaplan auf ihrem Gerät abgelegt haben und am Vorlesungsmitschrieb interessiert sind. Somit besitzt Annas Einheit zwei *potentielle Transaktionspartner*, mit denen eine beidseitig vorteilhafte Transaktion durchgeführt werden kann. Es liegt an Annas Einheit, einem dieser potentiellen Partner ihr Vertrauen auszusprechen und somit eine Transaktion einzugehen. In der Szenariobeschreibung wird die Entscheidung für Annas Einheit dadurch vereinfacht, dass sich nur Claude im Informatik-Bau, also in Reichweite befindet. Die Transaktion findet daher zwischen Annas und Claudes Einheiten statt. Weiterhin werden die ausgetauschten Informationen durch den Nutzen, den der Gegenüber durch den Erhalt der Information erfährt, und die Kosten, die die Bereitstellung mit sich bringt, charakterisiert. Auch dieses *Nutzen-/Kostenverhältnis* der Aktionsausführung geht in die Vertrauensentscheidung der Einheit Annas ein.

Das Simulative Kooperationsturnier nimmt diese Überlegungen in die folgende Modellierung der Kooperation auf: Im Laufe der Simulation werden wiederholt Transaktionsgelegenheiten er-

schaffen, die einer zufällige Einheit zugewiesen werden. Außerdem werden zufällig Einheiten bestimmt, die als potentielle Transaktionspartner in Frage kommen. Das Nutzen-/Kostenverhältnis der beiden auszuführenden Aktionen wird dabei ebenfalls bestimmt. Anschließend wird die Einheit, die die Transaktionsgelegenheit besitzt, informiert. Dabei ist den Einheiten nicht bewusst, wie viele Transaktionsgelegenheiten ihnen verbleiben. Dies entspricht der Situation des Campus-Szenarios, in dem keiner der Studenten genau weiß, wie oft er in Zukunft die Informationen Anderer noch benötigen wird. Allerdings besitzen Studenten mit Blick auf die von ihnen geplante Teilnahmedauer eine a priori Einschätzung der zu erwartenden Transaktionsgelegenheiten. Dies wird für die Bewertung und Zuweisung der Gegenstrategien in Abschnitt 10.1.2 berücksichtigt.

Das Simulative Kooperationsturnier lässt die Quantifizierung der Parameter des Kooperationsmodells offen. Sie werden in Abschnitt 10.2 variiert. Das Campus-Szenario gibt Aufschluss darüber, in welchen Größenordnungen sie sich bewegen: **(1)** Im Laufe der Teilnahme am Informationssystem besitzt eine Einheit im Schnitt sicherlich einige Transaktionsgelegenheiten. So fertigt zum Beispiel Anna fast täglich Vorlesungsmitschriebe an und der Mensaplan ist Claude oder David jeden Tag neu zugänglich. Hunderte von Transaktionsgelegenheiten sind für die Einheiten im Laufe ihrer Teilnahme am Informationssystem daher nicht ungewöhnlich. In der simulativen Evaluation gehen wir von einer konservativeren Schätzung aus, um zu berücksichtigen, dass unter Umständen nicht alle Einheiten derart intensiv am Informationssystem teilnehmen. **(2)** Die Zahl der potentiellen Transaktionspartner kann sehr unterschiedlich ausfallen. Den Mensaplan haben unter Umständen fast alle Geräte bei sich abgelegt. Vorlesungsmitschriebe sind hingegen sehr viel seltener vorhanden. In Folge dessen ist davon auszugehen, dass nur ein kleiner Teil der Einheiten potentielle Transaktionspartner sind. **(3)** Auch das Nutzen-/Kostenverhältnis ist von den einzelnen Aktionen abhängig. Ein Vorlesungsmitschrieb ist größer als ein Mensaplan und verursacht dadurch mehr Kosten für die Übermittlung. Ob ein Vorlesungsmitschrieb oder ein Mensaplan mehr Nutzen mit sich bringt hängt hingegen von der jeweiligen Situation des Studentens ab. Bei der Transaktion zwischen Claudes und Annas Gerät übersteigt gemäß der Szenariobeschreibung der Nutzen die Kosten um mehrere Größenordnungen. Beide Studenten benötigen die Information des jeweilig Anderen und empfinden die Kosten als minimal. In der simulativen Evaluation gehen wir dennoch von einer konservativeren Schätzung des Nutzen-/Kostenverhältnisses aus, um zu berücksichtigen, dass unter Umständen nicht alle Transaktionen einen derart hohen Mehrwert für die Benutzer mit sich bringen.

Modell der Kommunikation. Das in DIANEmu eingebaute Bewegungsmodell der Studenten beeinflusst nachhaltig die Möglichkeiten zur Kommunikation. Aufgrund der Bewegung der Einheiten untereinander kann es jederzeit zu Kommunikationsabbrüchen kommen. Die Übertragungsreichweite der Geräte wird im Simulativen Kooperationsturnier gemäß der Erfahrungen aus [JKSS04, ZEE03] auf 50 Meter eingestellt. Damit können die Einheiten zu jedem Zeitpunkt immer nur mit einem kleinen Teil der anderen Einheiten kommunizieren.

Die Wahrscheinlichkeit von Kommunikationsabbrüchen ist während der Aktionsausführung höher als beim Versenden einfacher Nachrichten. Dennoch haben wir in Abschnitt 1.2.2 gezeigt, dass selbst bei aufwändigen Aktionen die Wahrscheinlichkeit eines Kommunikationsabbruchs insgesamt sehr gering ist.

Einen weiteren Modellparameter stellen die Kosten für die Kommunikation dar. Sie sind weitaus geringer als die Kosten für die Aktionsausführung. Dies ergibt sich aus zweierlei Gründen: Einerseits sind einfache Nachrichten mehrere Größenordnungen kürzer als ausgetauschte Informationen wie zum Beispiel Vorlesungsmitschriebe. Andererseits stellt insbesondere das Ausführen von Informationsdiensten wie der Konversionsdienst auf Bobs Gerät einen weitaus höheren Auf-

wand dar als das Versenden einer kurzen Nachricht. Das geringe Verhältnis zwischen den Kosten der Nachrichtenübermittlung zu den Kosten der Aktionsausführung wird weiter nachhaltig vermindert, indem das zugrunde liegende Transportsystem Techniken wie das Piggy-Backing anwendet [RT99]. In diesem Fall können mehrere kurze Nachrichten zu einer Nachricht zusammengefasst werden, wodurch der Aufwand für das Versenden von Nachrichten weiter gesenkt wird.

Modell der Einheiten. In der Beschreibung des Campus-Szenarios kommen lediglich fünf Studenten vor. Es ist jedoch zu erwarten, dass insgesamt weitaus mehr Studenten am Informationssystem mit ihren Informationsgeräten teilnehmen. Eine Teilnehmerzahl von einigen hundert ist daher nicht unrealistisch. Dies entspricht in etwa auch den Möglichkeiten des Rahmenwerks DIANEmu. Die Simulation von hundert Einheiten ist dort problemlos möglich. Mit Hilfe der heutigen Simulationsrechner lassen sich auch einige hundert Einheiten simulieren. Allerdings wird die Simulation dadurch enorm zeitaufwändig[2]. Ausgangspunkt der simulativen Evaluation wird daher ein Informationssystem sein, an dem um die hundert Einheiten teilnehmen. Die Skalierbarkeit des Informationssystems wird zusätzlich in einigen dedizierten Versuchsreihen untersucht.

Ein wichtiger Parameter des Modells der Einheiten ist die Populationsstruktur. Hierunter fällt nicht nur der Anteil normativer Einheiten, dessen Variation gemäß Abschnitt 9.1.2 besondere Aufschlüsse ermöglicht. Zudem ist festzulegen, wie hoch der Anteil der strategischen Einheiten, die eine gewisse Gegenstrategie verfolgen, ist. Die Antwort hierauf wird durch die Antizipation der Manipulation gegeben. In der Durchführung der Evaluation des Gesamtsystems wird daher das Ergebnis des ersten Schrittes des Evaluationsprozesses an dieser Stelle zu berücksichtigen sein.

Ein weiterer Bestandteil des Modells der Einheiten ist die Verteilung der Manipulationskosten zwischen den menschlichen Benutzern. Im Simulativen Kooperationsturnier werden sie im Verhältnis zu den durchschnittlichen Kosten der Aktionsausführung angegeben. Damit bedeuten beispielsweise Manipulationskosten von 5, dass der entsprechende Student sich genau dann zur Manipulation entscheidet, wenn sich seine Einheit dadurch das Ausführen von 5 Aktionen (oder entsprechend mehr Aktionen wenn Betrugskosten anfallen) erspart. Wie hoch sind die Manipulationskosten einzuschätzen? Einerseits haben wir in Abschnitt 5.2 gesehen, dass es schwer wiegende Hindernisse zur Manipulation gibt. Andererseits entstehen die Kosten der Aktionsausführung im Campus-Szenario lediglich aus der sehr unwesentlichen Entladung der Batterie des Gerätes. Die Manipulationskosten sind damit um einige Größenordnungen höher als die Kosten der Aktionsausführung. Ein Problem, das für das Campus-Szenario spezifisch ist, ist jedoch, dass es sich bei den menschlichen Benutzern um Studenten handelt. Bei ihnen ist es durchaus denkbar, dass sie die technischen und rechtlichen Hindernisse zur Manipulation als weniger nachhaltig wahrnehmen. Insbesondere könnte für einige wenige manipulationsfreudige Studenten das Überwinden der technischen Hindernisse einen Reiz an sich ausüben. Bei der Modellierung der Manipulationskosten müssen wir daher davon ausgehen, dass sich einige Studenten schon bei einem geringen Vorteil gegenüber der originalen Systemsoftware zur Manipulation entscheiden. Die Verteilung der Manipulationskosten, auf die die simulative Evaluation beruht, wird daher auf einer äußerst vorsichtigen Schätzung beruhen.

[2]Dass die Simulation aufwändig ist, wird ersichtlich, wenn wir uns vor Augen führen, welche Teile die Simulation umfasst: Zum einen sind alle teilnehmenden Informationsgeräte zu emulieren. Zum anderen ist im Hintergrund die ständige Veränderung der Simulationsumgebung, wie zum Beispiel die Bewegung der menschlichen Prinzipale, zu simulieren.

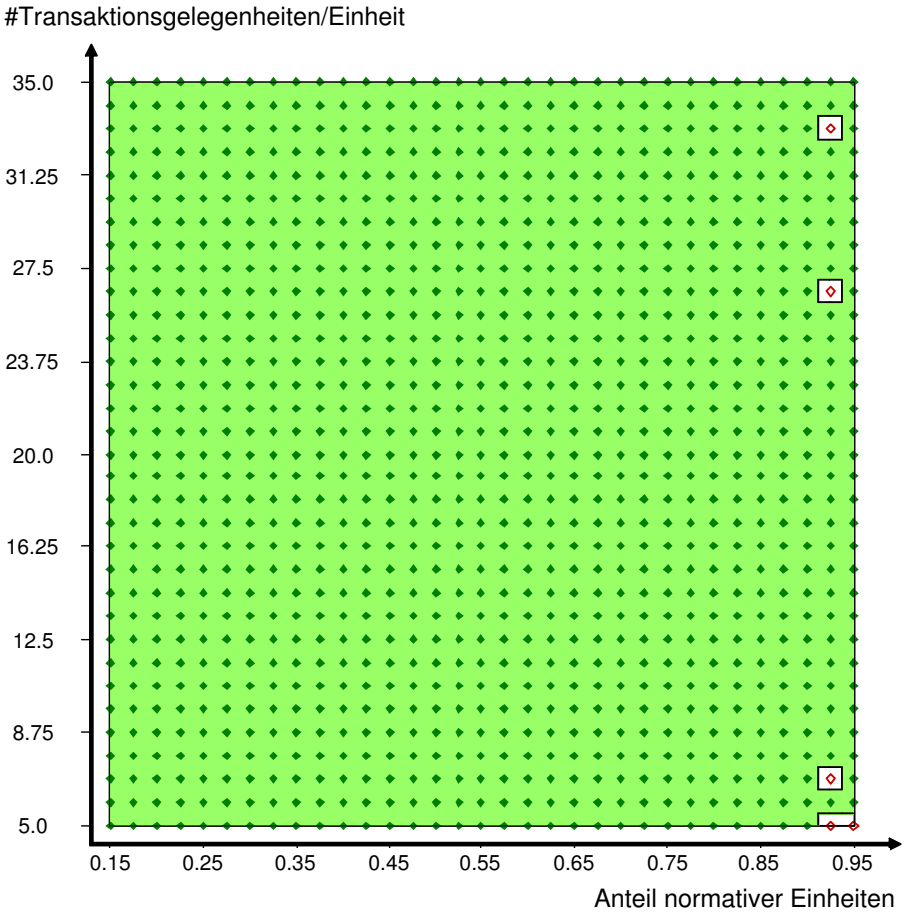

Abbildung 9.5: Beispielhaftes Ergebnis einer Sensibilitätsanalyse mit zwei variierten Dimensionen

9.2.2 Sensibilitätsanalyse

Laut Abschnitt 9.1.2 ist es für die Evaluation des Gesamtsystems unabdingbar, dass in mehreren Simulationsläufen einzelne Aspekte der Rahmenbedingungen variiert werden. In diesem Abschnitt erörtern wir, wie diese Anforderung im Simulativen Kooperationsturnier berücksichtigt wird.

Eine Erweiterung von DIANEmu um die Möglichkeit zu Sensibilitätsanalysen wird in [Zou05] durchgeführt[3]. Hierbei kann eingestellt werden, welche Teile der Rahmenbedingungen zu variieren sind. Für jeden diesen Teile ist dabei zu bestimmen, in welchem Bereich zu variieren ist und welche Schrittgröße zum Einsatz kommt. Die Sensibilitätsanalyse wird anschließend vollautomatisch durchgeführt.

Abbildung 9.5 zeigt ein beispielhaftes Ergebnis für die Sensibilitätsanalyse. Im Vergleich zur Abbildung 9.2 wird nicht nur der Anteil der normativen Einheiten im Bereich von 15% bis 95% in Schritten von 2,5% variiert. Zudem wird die durchschnittliche Zahl der Transaktionsgelegenheiten pro Einheit im Bereich von 5 bis 35 variiert. Die Interpretation der Punkte und Bereiche entspricht

[3]In der referenzierten Arbeit findet sich auch eine Methode zur automatisierten Optimierung der Entwurfsparameter. Diese Optimierung wird jedoch in der Gesamtevaluation des Kapitels 10 aufgrund des enormen Aufwands, den sie verursacht, nicht eingesetzt. Daher wird an dieser Stelle auf eine Einführung in diese Optimierungsmethode verzichtet.

der der eindimensionalen Sensibilitätsanalyse, die für Abbildung 9.2 in Abschnitt 9.1.2 besprochen worden ist. Ein weiteres Ergebnis der Sensibilitätsanalyse ist die Angabe, für welchen Bereich der Rahmenbedingungen das Informationssystem existenzfähig ist. Dieser Bereich wird wie bereits in Abbildung 9.2 eingefärbt. Hingegen ist der Bereich, in dem das Informationssystem nicht existenzfähig, weiß belassen.

Diese Art der Abbildung eignet sich der Natur nach nur für die Variation von bis zu zwei Dimensionen der Rahmenbedingungen. Die restlichen Dimensionen müssen fest vorgegeben werden, damit es zur Simulation kommen kann. Bei der Durchführung der Gesamtevaluation in Abschnitt 10.2 werden daher immer unterschiedliche Paare von Dimensionen variiert.

9.3 Interaktives Kooperationsturnier

Die Ergebnisse aus der Evaluation des Gesamtsystems sind nur dann aussagekräftig, wenn die Möglichkeiten zur Manipulation und dem daraus folgenden Betrugsverhalten auf eine realistische Weise berücksichtigt werden. Der Evaluationsprozess sieht hierfür als ersten Schritt das Finden von Gegenstrategien vor, die als Reaktion manipulationswilliger menschlicher Benutzer auf den normativen Systementwurf zu erwarten sind.

In diesem Abschnitt wird das Interaktive Kooperationsturnier als ein Werkzeug vorgestellt, mit dem solche Gegenstrategien gefunden werden können. Es geht von der Grundidee aus, dass viel versprechende Gegenstrategien von menschlichen Versuchspersonen im Zuge einer interaktiven Simulation gefunden werden. Die Grundidee und die Anforderungen, die sich aus ihr ergeben, werden in Abschnitt 9.3.1 besprochen. Abschnitt 9.3.2 zeigt anschließend, wie diese Anforderungen im Interaktiven Kooperationsturnier umgesetzt werden.

9.3.1 Grundidee und Anforderungen

Das Interaktive Kooperationsturnier ist für das systematische Finden von Gegenstrategien zuständig. In diesem Abschnitt gehen wir zunächst auf die Grundidee ein, wie dies zu bewerkstelligen ist. Anschließend werden aus der Grundidee einzelne Anforderungen abgeleitet, die das Interaktive Kooperationsturnier erfüllen muss, um die Grundidee effektiv umzusetzen.

Grundidee. Der Raum der Gegenstrategien ist groß, weil es eine Vielzahl von Möglichkeiten gibt, sich nicht an die Normen und Vorschriften des Systementwurfs zu halten. Damit sind wir nicht in der Lage, den Raum der Gegenstrategien exhaustiv zu durchsuchen. Auf der anderen Seite sind wir nur an solchen Gegenstrategien interessiert, deren Verfolgung zu einem besonders erfolgreichen Abschneiden führt. Da das Auftreten dieser viel versprechenden Gegenstrategien im realen System zu erwarten ist, handelt es sich bei ihnen um die Gegenstrategien, die zu finden sind.

Wie lässt sich der Raum der Gegenstrategien auf nicht exhaustive Weise nach eben diesen viel versprechenden Gegenstrategien durchsuchen? Gemäß Abschnitt 2.4 ist hierzu der Stand der Technik, dass der Systementwerfer selbst Angriffspunkte gegen seinen Entwurf durch eigenes Nachdenken identifiziert. Diese Vorgehensweise ist weder systematisch noch führt sie zu glaubwürdigen Evaluationsergebnissen. Die Grundidee des Interaktiven Kooperationsturniers besteht darin, dass die Gegenstrategien von Versuchspersonen gefunden werden. Dazu übernimmt jede Versuchsperson die Kontrolle über das Verhalten genau einer strategischen Einheit. Der Rest des Informationssystems und seine Rahmenbedingungen werden simuliert. Das Interaktive Kooperationsturnier ist dem Simulativen Kooperationsturnier daher ähnlich. Der Unterschied liegt

darin, dass im Interaktiven Kooperationsturnier Versuchspersonen am Simulationslauf teilnehmen und auf interaktive Weise das Verhalten ihrer Einheit bestimmen. Daraus erklärt sich der Name dieses Simulationswerkzeugs.

Versuchspersonen eignen sich für die Aufgabe, viel versprechende Gegenstrategien zu finden. Einerseits ähnelt die Kooperation im Informationssystem derjenigen in der menschlichen Gesellschaft, in der die Versuchspersonen ebenso Vertrauensentscheidungen treffen und kooperieren oder betrügen können. Andererseits sorgt das Interaktive Kooperationsturnier dafür, dass von den Versuchspersonen für die Teilnahme keine besonderen Vorkenntnisse oder abstraktes Denken verlangt werden. Die Versuchspersonen entwickeln spielerisch ihre jeweilige Strategie und müssen sich nicht a priori auf sie festlegen. Das Interaktive Kooperationsturnier ist dafür zuständig, dass die Versuchspersonen dabei auch tatsächlich viel versprechende Gegenstrategien entwickeln. Mit den Anforderungen, die sich daraus ergeben, befassen wir uns im nachfolgenden Paragraphen.

Anforderungen. Zunächst benötigt das Interaktive Kooperationsturnier eine *technische Grundlage*. Sie muss gewährleisten, dass einerseits Versuchspersonen interaktiv teilnehmen können und andererseits das Informationssystem und seine Rahmenbedingungen simuliert werden. Diese Simulation muss genauso realitätsnah wie die des Simulativen Kooperationsturniers sein. Abstriche am Realitätsgrad können aber nötig sein, um die anderen Anforderungen an das Interaktive Kooperationsturnier erfüllen zu können. Dies ist vertretbar, da die gefundenen Gegenstrategien im zweiten Schritt des Evaluationsprozesses durch eine Nachsimulation im Simulativen Kooperationsturnier unter realistischen Bedingungen bewertet werden. Eine weitere Anforderung an das Interaktive Kooperationsturnier ist, dass es das Verhalten einer jeden Einheit steuern kann. Diese Notwendigkeit ergibt sich bei normativen Einheiten direkt, da Versuchspersonen immer nur strategische Einheiten übernehmen. Allerdings können wir uns nicht darauf verlassen, dass es bei jedem Versuchslauf ausreichend Versuchspersonen gibt, damit alle strategischen Einheiten von je einer Versuchsperson gesteuert wird. Daher muss die Simulationsumgebung neben normativen Einheiten auch strategische Einheiten übernehmen können. Um zu einem realistischen Verhalten dieser automatisch gesteuerten strategischen Einheiten zu kommen, sind Gegenstrategien einzubeziehen, die bereits in früheren Versuchsläufen als viel versprechend identifiziert worden sind.

Eine Reihe weiterer Punkte ergeben sich aus der Forderung nach der *Zugänglichkeit und Erlernbarkeit* des Interaktiven Kooperationsturniers. Diese ist nötig, um den Kreis potentieller Versuchspersonen so weit wie möglich zu fassen und damit eine breite Grundlage für das Finden von Gegenstrategien zu erhalten. Prinzipiell soll jede interessierte Person in der Lage sein, am Kooperationsturnier teilzunehmen. Da sich solche Personen nicht notwendigerweise am selben Ort befinden, muss eine entfernte Teilnahme am Interaktiven Kooperationsturnier möglich sein. Außerdem dürfen von den Versuchspersonen keine besonderen Vorkenntnisse im Bereich der Informatik oder der Programmierung verlangt werden. Nur so ist sichergestellt, dass Personen (etwa aus dem Bereich der Soziologie) teilnehmen können, die zwar keine technischen Vorkenntnisse besitzen, aber für das spielerische Finden von Gegenstrategien geeignet sind. Ein besonderes Augenmerk muss dem Problem gewidmet werden, dass anfänglich keine der Versuchspersonen mit dem Entwurf der verteilten Vertrauensbildung aus Teil II dieser Arbeit vertraut ist. Es bedarf daher einer Methode, mit der dieser Entwurf den Versuchspersonen vermittelt wird, ohne sie zu überfordern oder ihr Interesse an der Teilnahme zu vermindern.

Die *Bedienbarkeit* des Interaktiven Kooperationsturniers stellt eine weitere Anforderung dar. Dabei ist zu gewährleisten, dass im Zuge der Interaktion mit dem Kooperationsturnier jede Versuchsperson ihren Willen bezüglich des Verhaltens ihrer Einheit durchsetzen kann. Hierzu müssen die Vorgänge des simulierten Gesamtsystems graphisch entsprechend aufbereitet werden, so dass

die Versuchspersonen eine Grundlage für das Treffen ihrer Entscheidungen erhalten. Außerdem müssen diese Entscheidungen auf einfache und unmissverständliche Weise dem Kooperationsturnier gegenüber vermittelt werden können.

Weiterhin ist es unumgänglich, dass die Versuchspersonen zur erfolgreichen Teilnahme und damit zum Entwickeln viel versprechender Gegenstrategien *motiviert* werden. Dass dies keineswegs selbstverständlich ist, zeigt die folgende Überlegung: Im Allgemeinen wird der Ehrgeiz einer Versuchsperson dadurch angestachelt, möglichst besser als die anderen Versuchspersonen abzuschneiden. Wenn eine Versuchsperson also vermutet, dass eine andere Versuchsperson ihr nahe kommt, so könnte sie versucht sein, dieser anderen Versuchsperson durch entsprechendes Verhalten ihrer eigenen Einheit zu schaden. Diese an sich selbstverständliche Verhaltensweise gefährdet jedoch den Erfolg des Interaktiven Kooperationsturniers. Im Gegensatz zu den menschlichen Prinzipalen eines realen Informationssystems, orientieren dann nämlich die Versuchspersonen ihr Verhalten nicht an der Maximierung ihres jeweiligen Individualnutzens. Als Ergebnis erhalten wir Gegenstrategien, die im Sinne der Nutzenmaximierung nicht viel versprechend sind. Für das Interaktive Kooperationsturnier muss daher gefordert werden, dass die Versuchspersonen ihr Verhalten an der Maximierung ihres Individualnutzens ausrichten. Da dies dem Naturel menschlicher Spieler unter Umständen widerspricht, sind entsprechende Vorkehrungen im Interaktiven Kooperationsturnier zu treffen.

Werden die oben genannten Anforderungen erfüllt, so eignet sich das Interaktive Kooperationsturnier zum Finden von Gegenstrategien. Dass dem so ist, zeigt die zusammenfassende Auflistung der Anforderungen:

- Das spielerische Entwickeln von Gegenstrategien ist möglich (technische Grundlage).
- Die Versuchspersonen wollen viel versprechende Gegenstrategien finden (Motivation).
- Die Versuchspersonen sind jederzeit in der Lage, ihre Intentionen umzusetzen (Bedienbarkeit).
- Eine breite Vielfalt von Gegenstrategien wird gefunden (Zugänglichkeit und Erlernbarkeit).

9.3.2 Umsetzung

Anhand der identifizierten Anforderungen wurde das Interaktive Kooperationsturnier als Simulationswerkzeug entwickelt. In diesem Abschnitt zeigen wir, wie darin die einzelnen Anforderungen umgesetzt werden. Weitere Details zur Entwicklung des Interaktiven Kooperationsturniers finden sich in [Fäh05].

Die nachfolgende Besprechung orientiert sich größtenteils an den zwei Darstellungen des Interaktiven Kooperationsturniers aus Abbildung 9.6 und 9.7. An den einzelnen Teilen der Benutzungsoberfläche des Interaktiven Kooperationsturniers wird die Umsetzung der Anforderungen aufgezeigt.

Technische Grundlage. Das Interaktive Kooperationsturnier wurde als Web-Applikation umgesetzt. Dazu wurde nicht nur die Implementierung des eigenen Entwurfs aus dem Simulativen Kooperationsturnier übernommen und entsprechend angepasst.

Das Interaktive Kooperationsturnier modelliert die Rahmenbedingungen der Simulation in Anlehnung an das Simulative Kooperationsturnier. Durch die Berücksichtigung der Anforderungen an die Bedienbarkeit mussten zwei Aspekte der Rahmenbedingungen im Interaktiven Kooperationsturnier angepasst werden:

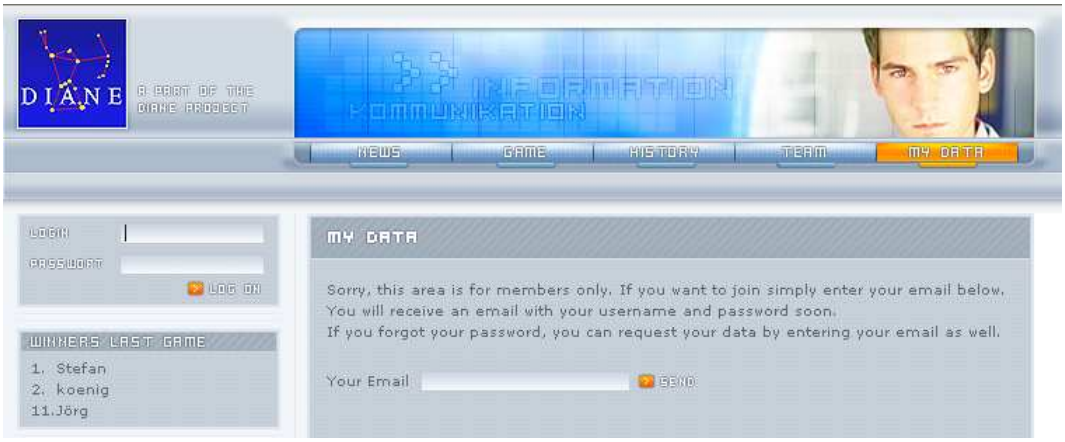

Abbildung 9.6: Portal des Interaktiven Kooperationsturniers

- *Zeitliche Abfolge:* Ein grundsätzliches Problem stellt das Zusammenspiel zwischen Simulation und Versuchspersonen dar. Die Versuchspersonen benötigen für das Treffen und Vermitteln ihrer Entscheidungen Zeit. Zudem werden sie unter Umständen im Versuchsverlauf von ihrer Umgebung gestört und können für eine kurze Zeit keine Entscheidungen treffen. Diese Einschränkungen sind im Interaktiven Kooperationsturnier berücksichtigt worden. Dazu wurde zum einen eine Orientierung des Versuchsablaufs an *Runden* eingeführt. Jede Einheit erhält nur am Anfang einer Runde Transaktionsgelegenheiten. Durch Bekanntgabe der Rundendauer weiß jede Versuchsperson, zu welchem Zeitpunkt ihre besondere Aufmerksamkeit gefordert ist. Außerdem finden Änderungen im Modell der Kommunikation nur am Ende einer Runde statt. Konkret bedeutet dies, dass die Erreichbarkeit der Einheiten untereinander für die Dauer einer Runde unverändert bleibt. Dadurch wird gewährleistet, dass länger überlegende oder für kurze Zeit verhinderte Versuchspersonen keinen Nachteil erfahren. Eine weitere Anpassung betrifft die Asynchronität der Kommunikation zwischen den Einheiten. Da eine Versuchsperson im Allgemeinen nicht in der Lage ist, instantan auf Anfragen zu Transaktionen, Empfehlungen oder Bürgschaftsbeziehungen zu antworten, wird ihr hierfür die Dauer einer Runde Zeit gegeben. Dadurch wird auch erreicht, dass sich die Versuchspersonen für ihre Entscheidungen ausreichend Zeit nehmen.

- *Transaktionsmodell:* Der Verständlichkeit wegen ist das Transaktionsmodell vereinfacht. Dabei wurde die beidseitige Aktionsausführung der an einer Transaktion beteiligten Einheiten aufgegeben. Stattdessen führt nur eine Einheit eine Aktion im eigentlichen Sinne aus. Die andere Einheit überweist ihr dafür Individualnutzen in Form einer virtuellen Währung. Dieses Transaktionsmodell wurde gewählt, weil es den Versuchspersonen intuitiver erscheint. Es lässt sich jedoch als eine Sicht auf das ursprüngliche realistische Transaktionsmodell interpretieren: Die Übergabe des Individualnutzens stellt das Ausführen einer Aktion dar, deren den Nutzen-/Kostenverhältnis gleich eins ist. Somit ergibt sich der gesamte Mehrwert einer Transaktion aus dem Ausführen der anderen Aktion. Dies wird im Interaktiven Kooperationsturnier dadurch berücksichtigt, dass das Nutzen-/Kostenverhältnis dieser zweiten Aktion im Vergleich zum Simulativen Kooperationsturnier entsprechend höher eingestellt ist.

Die Implementierung des Interaktiven Kooperationsturniers ist in der Lage, sowohl normative als auch strategische Einheiten zu steuern. Die Durchführung der Versuchsreihen wird von einem

Abbildung 9.7: Momentaufnahme des Interaktiven Kooperationsturniers

Administrator geleitet. Er ist auch dafür zuständig, bereits gefundene Gegenstrategien algorithmisch festzuhalten und im Interaktiven und Simulativen Kooperationsturnier zu implementieren.

Zugänglichkeit und Erlernbarkeit. Das Interaktive Kooperationsturnier ist als Web-Applikation mit graphischer Benutzungsschnittstelle implementiert. Damit sind Versuchspersonen an keinen bestimmten Ort gebunden, um am Kooperationsturnier teilzunehmen. Ihnen genügt hierfür ein gewöhnlicher Internet Browser, der heutzutage auf jedem Desktop Computer anzutreffen ist, und Zugang zum Internet. Insbesondere sind für die Teilnahme also nicht die Installation einer bestimmten Software oder besondere Hardware-Voraussetzungen notwendig. Damit ist der Kreis potentieller Versuchspersonen sehr weit gefasst. Zur Teilnahme ist lediglich eine Anmeldung auf dem Portal des Interaktiven Kooperationsturniers erforderlich. Abbildung 9.6 zeigt, dass hierfür keine persönlichen Daten übermittelt werden müssen. Das Angeben der E-Mail Adresse reicht zu diesem Zweck aus. Damit ist die Anforderung erfüllt, dass das Interaktive Kooperationsturnier allen potentiellen Versuchspersonen *zugänglich* ist. Die *Erlernbarkeit* des Interaktiven Kooperationsturniers wird durch eine Reihe von Maßnahmen hergestellt, die wir im Folgenden erörtern.

Es gibt drei generelle Orientierungshilfen für die Versuchspersonen. **(1)** Das *Glossar* bespricht die Terminologie des Kooperationsturniers und zeigt die Konzepte auf, die ihm zugrunde liegen. An dieser Stelle erfährt eine Versuchsperson von den Details des Entwurfs der verteilten Vertrauensbildung. Dieses Glossar ist um einen Abschnitt zu häufig gestellten Fragen (engl.: frequently asked questions, kurz *FAQ*) ergänzt, das vom Administrator des Interaktiven Kooperationsturnier ständig ergänzt wird. Dadurch wird die Orientierungshilfe vor allem bei der ersten Teilnahme einer Versuchsperson weiter verstärkt. Sowohl das Glossar als auch der FAQ-Abschnitt ist jederzeit

über eine Leiste, die in der Momentaufnahme der Abbildung 9.7 links dargestellt ist, erreichbar. **(2)** Der Zugang zu Hintergrundinformationen wird auch im Hauptbereich der Benutzungsschnittstelle gewährleistet, der sich rechts der Leiste befindet. Für jeden Begriff, der mit einem Konzept verbunden ist, wird dort ein *Fragezeichen* dargestellt, über das Hintergrundinformationen abgerufen werden können. So gibt es in der Momentaufnahme zum Beispiel Fragezeichen zu den Themengebieten Transaktionsgelegenheiten und zu den Möglichkeiten des Transaktionsverhaltens. **(3)** Zusätzlich sorgt ein *Forum* dafür, dass sich die Versuchspersonen auch untereinander über das Interaktive Kooperationsturnier austauschen können. Hierbei ist darauf aufzupassen, dass sich die Versuchspersonen nicht untereinander über ihr Verhalten im Kooperationsturnier absprechen. Dies würde insbesondere den Empfehlungsmechanismus des Entwurfs der verteilten Vertrauensbildung umgehen und dadurch zu verfälschten Ergebnissen führen. Daher setzt der Administrator im Forum die Verfahrensweise durch, dass während eines Simulationslaufes die Versuchspersonen nicht ihre im Turnier gemachten Erfahrungen austauschen dürfen.

Die generellen Orientierungshilfen sorgen dafür, dass Versuchspersonen, die Informationen über das Interaktive Kooperationsturnier benötigen, diese auch finden. Dies reicht technisch weniger versierten Versuchspersonen bei der ersten Teilnahme am Turnier jedoch nicht aus. Zusätzlich bedürfen sie einen Mechanismus, mit dem sie sich langsam in die vielfältigen Bereiche vorarbeiten können, in denen Entscheidungen zu treffen sind. So kann von einer Versuchsperson anfänglich kaum erwartet werden, dass sie sich der Folgen ihres Empfehlungs- und Bürgschaftsverhaltens bewusst ist. Allerdings können ihr durchaus Entscheidungen zum Eingehen von Transaktionen und zum Verhalten darin zugetraut werden. Daher sollten Versuchspersonen anfangs nur das Transaktionsverhalten festlegen müssen. Es bedarf also einer Möglichkeit, mit der das Interaktive Kooperationsturnier bei Bedarf gewisse Arten von Entscheidungen der Versuchsperson abnimmt. Diese Überlegung findet im Interaktiven Kooperationsturnier durch die Implementierung einer *Computer-Hilfe* Berücksichtigung. Über die Leiste lassen sich die Einstellungen (engl.: settings) abändern, mit denen eine Versuchsperson am Turnier teilnimmt. Genauer gesagt kann sie auswählen, für welche Bereiche sie das Verhalten ihrer Einheit bestimmen will. Für die anderen Bereiche lässt sich die Strategie festlegen, die die eigene Einheit verfolgt. So lässt sich zum Beispiel für das Empfehlungsverhalten einstellen, ob es den Vorschriften des Systementwurfs folgt oder aktiver oder passiver als dieses ist. Diese Teilstrategien sind bewusst einfach gehalten, damit die Versuchspersonen dazu gebracht werden, schließlich alle Bereiche des Verhaltens ihrer Einheit selbst zu bestimmen.

Bedienbarkeit. Grundlage für die Bedienbarkeit des Interaktiven Kooperationsturniers ist die Entwurfsentscheidung, dass so viele Sachverhalte wie möglich visuell dargestellt werden und die Interaktion keine Tastatureingaben erfordert. Im Folgenden erörtern wir, wie sich diese Entscheidung im Einzelnen auf die Entwicklung des Interaktiven Kooperationsturniers auswirkt.

Jede Einheit wird mit dem Namen einer berühmten Persönlichkeit versehen. Dieser Name dient ihr im Turnierverlauf als Identifikator. Zusätzlich zum Namen erscheint ein *Avatar* eben dieser Persönlichkeit, um die Wiedererkennung einer Einheit weiter zu erleichtern. In der Momentaufnahme der Abbildung 9.7 steuert die Versuchsperson zum Beispiel eine Einheit, die den Identifikator SCHMELING besitzt. Bei den Transaktionsgelegenheiten werden jeweils die Namen und Avatare der potentiellen Transaktionspartner dargestellt. Bei der Zuweisung der Identifikatoren zu den Einheiten kommen zwei Aspekte zur Berücksichtigung:

- Die berühmten Persönlichkeiten sind aus Domänen wählen, in denen im Vorhinein keine besonders starken Zuneigungen der Versuchspersonen bestehen. So scheiden zum Beispiel

Politiker aus, da bei ihnen eine Verzerrung des Verhaltens der Versuchspersonen zu erwarten ist. Aus demselben Grund sind alle gewählten Persönlichkeiten männlich. Insgesamt umfasst das Interaktive Kooperationsturnier über 200 Persönlichkeiten, die aus den Bereichen Sport, Kunst, Musik, Film, Geschichte und Wissenschaft stammen.

- Der Identifikator einer Einheit darf keinen Aufschluss über ihren Typ geben. Daher wird jeder Einheit vor jedem Turnier zufällig eine Persönlichkeit zugeordnet. Aus der Sicht einer Versuchsperson hat dies zur Folge, dass sie bei jedem Turnier unter einem anderen Namen teilnimmt.

Das an sich abstrakte Konzept des Individualnutzens wird mit Hilfe der *Metapher der Kalorien* illustriert. Der Transaktionskontext wird gemäß dieser Metapher mit dem Symbol einer Mahlzeit dargestellt, deren Kaloriengehalt die Höhe des Transaktionskontextes widerspiegelt. Das Interaktive Kooperationsturnier unterscheidet acht solcher Mahlzeiten, die vom Apfel zu einer Pizza reichen. In der Momentaufnahme sind zum Beispiel die Transaktionsgelegenheiten jeweils mit einem Kontext versehen, der mit einem Symbol für einen Hamburger, Pommes Frites und einem Eis illustriert wird. Hierdurch sind Versuchspersonen in der Lage, sich intuitiv ein Bild von der Wichtigkeit der Transaktion zu machen. Zu der Metapher der Kalorien passt auch die Darstellung des Individualnutzens, der mit einem Symbol für eine Gabel versehen ist.

Die Informationen, die der Einheit einer Versuchsperson zugänglich sind, werden im *Verhaltensprotokoll* (engl.: track record) zusammengefasst. Die Darstellung gibt bezüglich jeder Einheit Auskunft über ihre Bürgen, den Ausgang bisheriger Transaktionen mit ihr und die Empfehlungen, die bisher über sie ausgestellt worden sind. Damit die Versuchspersonen nicht bei jeder Entscheidungssituation auf das Verhaltensprotokoll zurückgreifen müssen, sind die wichtigsten Informationen über eine Einheit jederzeit in aggregierter Form als Symbole dargestellt. Die Momentaufnahme umfasst drei Arten solcher Symbole: **(1)** Die Handschellen zeigen an, dass die Einheit namens KAFKA die Einheit der Versuchsperson zuvor betrogen hat. **(2)** Das Ausrufezeichen bei KAFKA oder BLOM bedeutet, dass diese Einheiten in Vergangenheit von der Versuchsperson betrogen worden sind. **(3)** Das Symbol des Händeschüttelns gibt an, dass für BLOM und FAULKNER Bürgen bekannt sind. Außer den Symbolen gibt es eine weitere Form der Informationsaufbereitung: Durch das Anklicken des Avatars einer Einheit wird der Versuchsperson der Typglauben angezeigt, den sie gemäß ihrer Informationslage über diese Einheit haben müsste.

Das Verhaltensprotokoll ist um einen *Ticker* ergänzt, in dem die Ereignisse der letzten Runden aus der Sicht der Einheit der Versuchsperson zusammengefasst sind. Damit wird sichergestellt, dass die Versuchsperson über die Vorgänge im Informationssystem unterrichtet wird. In der Momentaufnahme befindet sich der Ticker am unteren Bildrand.

Das Interaktive Kooperationsturnier sorgt zudem dafür, dass die Versuchspersonen über die Rahmenbedingungen (wie etwa die Wahrscheinlichkeit für Kommunikationsabbrüche) informiert sind. Diese Rahmenbedingungen sind auf der *Übersichtsseite* des Turniers jederzeit abrufbar. Zudem ist der Administrator des Interaktiven Kooperationsturniers für das regelmäßige Versenden einer *Newsletter* zuständig, in der die Rahmenbedingungen bevorstehender Turniere vorgestellt werden.

Motivation. Das Interaktive Kooperationsturnier erreicht durch eine Reihe von Maßnahmen, dass die Versuchspersonen ihr Verhalten an der Maximierung ihres Individualnutzens ausrichten. Diese Maßnahmen werden im Folgenden besprochen.

Der Individualnutzen, den eine Versuchsperson im Laufe des bisherigen Turniers erhalten hat, wird jederzeit unter der Leiste der Benutzungsoberfläche angezeigt. Entsprechend der Metapher

der Kalorien ist die Zahl mit einem Gabelsymbol versehen. Dennoch reicht dies unter Umständen nicht aus, damit die Versuchspersonen einen intuitiven Zugang zum Grad ihres Erfolgs erhalten. Daher wird zusätzlich die Höhe des Individualnutzens mit Hilfe eines *Statusmännchens* visualisiert. Das Interaktive Kooperationsturnier hält fünf solcher Statusmännchen vor, das je nach Kalorienzahl dicker oder dünner ist. Wie entscheidet sich, welches der Statusmännchen angezeigt wird? Auf den ersten Blick erscheint es sinnvoll, dass das Statusmännchen entsprechend der momentanen Platzierung der Versuchsperson angezeigt wird. Dadurch würde jede Versuchsperson jedoch möglichst besser als die anderen Versuchspersonen abzuschneiden wollen. Damit würde unser Ziel unterlaufen, dass die Versuchspersonen ihr Verhalten allein an der Maximierung ihres Individualnutzens ausrichten. Daher wird im Interaktiven Kooperationsturnier das Statusmännchen abhängig von der absoluten Leistung der Versuchsperson angezeigt. Dazu wird berechnet, welcher Individualnutzen von einer guten, mittelmäßigen oder schlechten Versuchsperson in der bisherigen Rundenzahl zu erwarten ist. Durch Vergleich mit dem aktuellen Individualnutzen erfolgen eine Einordnung der Versuchsperson und die Anzeige eines entsprechenden Statusmännchens. In der Momentaufnahme der Abbildung 9.7 hat zum Beispiel die Versuchsperson recht gut abgeschnitten, weswegen das Statusmännchen relativ dick ist.

Am Ende eines Turniers ist es möglich, die relative Platzierung der Versuchspersonen untereinander in einer *Bestenliste* mitzuteilen, da die Versuchspersonen ihr Verhalten nicht mehr nachträglich anhand dieser Information revidieren können. Es gibt zwei Arten der Bestenlisten: Eine Bestenliste des letzten ausgetragenen Turniers wird stets auf dem Portal in der Leiste angezeigt. Hierbei sind nur die drei Versuchspersonen, die am besten abgeschnitten haben, und ihre relative Platzierung auch gegenüber der anderen Einheiten aufgelistet. Die Darstellung des Portals in Abbildung 9.6 zeigt zum Beispiel, dass zwei Versuchspersonen die erfolgreichsten Einheiten des letzten Turniers gesteuert haben. Vor der Einheit der drittbesten Versuchsperson liegen jedoch weitere Einheiten, die vom Interaktiven Kooperationsturnier gesteuert wurden. Durch diese Art der permanent abrufbaren Bestenliste wird die Motivation der Versuchspersonen zur erfolgreichen Teilnahme am Turnier weiter erhöht. Eine zweite Art der Bestenliste ist in Abbildung 9.8 dargestellt. Sie wird am Ende eines jeden Turniers den Versuchspersonen mitgeteilt. In dieser Bestenliste befinden sich nicht nur Informationen über die Platzierung der automatisch gesteuerten Einheiten sondern auch Hintergründe für das jeweilige Abschneiden der Einheiten. Dabei können die Versuchspersonen auch ganz gezielt die Momentaufnahmen der Bestenliste für einen bestimmten Zeitpunkt des Turniers abrufen. Im Einzelnen sind für jede Einheit folgende Informationen verfügbar:

- Wurde die Einheit von einer Versuchsperson gesteuert? Wenn sie automatisch gesteuert wurde, war sie normativ oder strategisch? Welche Gegenstrategie kam bei einer strategischen Einheit zum Einsatz? In der Abbildung belegen Versuchspersonen den ersten und dritten Rang. Zwei normative Einheiten sind jeweils Zweiter und Vierter. Eine strategische Einheit mit der Gegenstrategie namens NormativeRetaliator ist Fünfter.

- Wie hat die Einheit insgesamt von ihrem Individualnutzen her abgeschnitten? Wie auch während des Turniers wird der Individualnutzen durch ein Statusmännchen visualisiert.

- Wie sind die Kennzahlen für das Abschneiden der Einheit? Aufgelistet sind der Individualnutzen aus der Teilnahme an Transaktionen, das eigene Transaktionsverhalten und das des jeweiligen Transaktionspartners, die Häufigkeit unbeabsichtigten Betrugsverhaltens und die Zahl der Bürgen.

Place	User	Strategy	Status	Total Calories	Gains through buying	Gains through selling	Won through betrayls	Lost through betrayls	Lost through Noise	Total Bails ever
1	Verdi	human		912.3	22 520.77	0 0	27 489.9	1 7.91	6 82.26	0
2	Ford	normative		910.4	27 736.93	30 259.05	0 0	1 24.55	5 55.48	9
3	Blom	human		878.22	25 416.72	28 274.78	16 297.78	4 48.82	3 65.23	6
4	Ravel	normative		798.01	22 609.61	25 261.3	0 0	3 56.62	2 24.55	3
5	Woods	normative_retaliator		787.25	25 563.76	29 291.59	0 0	0 0	3 128.33	6

Abbildung 9.8: Ausschnitt aus der Bestenliste eines Interaktiven Kooperationsturniers

Durch diese detaillierte Darstellung erhalten die Versuchspersonen Aufschluss über den Zusammenhang zwischen ihrem eigenen Verhalten und den Grad ihres Erfolgs. In nachfolgenden Turnieren sind die Versuchspersonen dadurch in der Lage, ihr Strategie entsprechend anzupassen. Die detaillierte Bestenliste ist also dafür zuständig, dass das Verhalten der Versuchspersonen über verschiedene Turniere hinweg rückgekoppelt ist und dadurch immer erfolgreichere Gegenstrategien gefunden werden.

Eine weitere Motivationsquelle stellt die *Befragung* der erfolgreichsten Versuchspersonen dar. Sie wird vom Administrator des Interaktiven Kooperationsturniers nach der Auswertung der Turnierergebnisse durchgeführt. Dabei wird im Gespräch die Gegenstrategie identifiziert, die die Versuchsperson spielerisch entwickelt und erfolgreich angewendet hat. Dem Administrator steht hierfür neben den Aussagen der Versuchsperson auch ein detailliertes Protokoll des Turniers zur Verfügung, das vom Interaktiven Kooperationsturnier festgehalten und bereitgestellt wird. Die Erfahrung zeigt, dass die Versuchspersonen gerne dazu bereit sind, ihr Strategie dem Administrator gegenüber zu erklären. Die Motivation dazu ergibt sich zum Teil aus der Ehrung, die der Versuchsperson durch die Nachfrage des Administrators zuteil wird. Es werden nämlich nur die Versuchspersonen nachgefragt, die besonders gut abgeschnitten haben. Insgesamt erreichen wir dadurch, dass die viel versprechenden Gegenstrategien, die im Interaktiven Kooperationsturnier spielerisch gefunden werden, formuliert werden und für die weiteren Schritte des Evaluationsprozesses zur Verfügung stehen.

9.4 Zusammenfassung

Ein abschließendes Urteil zur Existenzfähigkeit des Informationssystems kann nur im Zuge einer umfassenden Evaluation erfolgen. Hierfür wurde in diesem Kapitel eine *Methodik* entwickelt. Sie zieht die Simulation als das wesentliche Mittel der Evaluation heran. In diesem Zusammenhang

haben wir die entscheidenden Vorteile der Simulation gegenüber der Analyse aufgezeigt. In die simulative Evaluation gehen drei Elemente ein: Es sind die Rahmenbedingungen (die Benchmark), der Entwurf der verteilten Vertrauensbildung (der Entwurfspunkt) und die Metriken. Letztere ergeben sich aus der Ausrichtung der Evaluation an unserem Ziel, die These von der Existenzfähigkeit eines Informationssystems wie dasjenige des Campus-Szenarios zu verifizieren. Es wurde ein Kriterium für die Existenzfähigkeit und eine Maßzahl für die Güte des Entwurfs abgeleitet. Es hat sich gezeigt, dass eine ganze Reihe von Simulationsläufen vonnöten sind, um die Abhängigkeit der Erfüllung des Kriteriums von den Eigenheiten der Rahmenbedingungen abschätzen zu können. Dies führte uns zu Formulierung eines zweistufigen Evaluationsprozesses. Die Antizipation der Manipulation bildet dabei die Voraussetzung für die Evaluation des Gesamtsystems. Der Evaluationsprozess sieht zwei Simulationswerkzeuge vor, mit denen wir uns im Kapitel beschäftigt haben.

Das *Simulative Kooperationsturnier* ist für die Evaluation des Gesamtsystems zuständig. Als Grundlage wurde das Rahmenwerk DIANEmu gewählt, da in ihm das Bewegungsmodell menschlicher Benutzer für das Campus-Szenario bereits eingebaut ist. In diesem Rahmenwerk wurde der eigene Ansatz implementiert. Für die Rahmenbedingungen des Campus-Szenarios, die in DIANEmu keine Berücksichtigung finden, wurden entsprechende Modelle entwickelt, die eine realitätsnahe Simulation ermöglichen. Außerdem wurde das Simulationswerkzeug um die Möglichkeit zu Sensibilitätsanalysen erweitert, da die Variation einzelner Aspekte der Rahmenbedingungen für die Evaluation des Gesamtsystems unabdingbar ist.

Die Ergebnisse aus der Evaluation des Gesamtsystems sind nur dann aussagekräftig, wenn die Möglichkeiten zur Manipulation und dem daraus folgenden Betrugsverhalten auf eine realistische Weise berücksichtigt werden. Daher sieht der Evaluationsprozess als ersten Schritt das Finden von Gegenstrategien vor, die als Reaktion manipulationswilliger menschlicher Benutzer auf den normativen Systementwurf zu erwarten sind. Hierbei kommt das *Interaktive Kooperationsturnier* zum Einsatz. Es geht von der Grundidee aus, dass viel versprechende Gegenstrategien von menschlichen Versuchspersonen im Zuge einer interaktiven Simulation gefunden werden. Voraussetzung hierfür ist die Erlernbarkeit und Bedienbarkeit durch die Versuchspersonen sowie ihre Motivation zur erfolgreichen Teilnahme. Es wurde gezeigt, wie im Interaktiven Kooperationsturnier diese Anforderungen umgesetzt werden.

Kapitel 10

仁者安仁 智者利仁

"Der Gute begnügt sich mit der Güte. Hingegen verfolgt jemand, der lediglich weise ist, die Güte in dem Glauben, dass sich dies für ihn auszahlt."
(Gespräche und Aussprüche des Konfuzius, 4.2)

Durchführung der Evaluation

Der Entwurf der verteilten Vertrauensbildung zielt auf die Existenzfähigkeit von Informationssystemen wie dasjenige des Campus-Szenarios. In der simulativen Evaluation des Entwurfs ist herauszufinden, unter welchen Rahmenbedingungen diese Existenzfähigkeit gewährleistet ist. Zu diesem Zweck wurde im vorigen Kapitel eine Methodik entwickelt, die den Evaluationsprozess festlegt und für die notwendige Unterstützung durch Simulationswerkzeuge sorgt.

In diesem Kapitel setzen wir diese Methodik ein, um die verteilte Vertrauensbildung zu evaluieren. Unser Vorgehen orientiert sich dabei an den Schritten des Evaluationsprozesses: In Abschnitt 10.1 antizipieren wir basierend auf dem Interaktiven Kooperationsturnier, in welche Richtungen die Manipulation der originalen Systemsoftware zu erwarten ist. Anschließend wird in Abschnitt 10.2 die eigentliche Evaluation des Gesamtsystems durchgeführt. Durch den Einsatz des Simulativen Kooperationsturniers erwarten wir uns dabei Aufschluss darüber, unter welchen Rahmenbedingungen unsere These von der Existenzfähigkeit des Informationssystems zutrifft.

10.1 Antizipation der Manipulation

Im ersten Schritt des Evaluationsprozesses geht es darum zu antizipieren, wie manipulationswillige menschliche Benutzer auf den normativen Systementwurf reagieren. Diese Antizipation der Manipulation ermöglicht es, dass in der Evaluation des Gesamtsystems etwaiges Betrugsverhalten strategischer Einheiten auf eine realistische Weise berücksichtigt wird.

Bei der Durchführung dieser Antizipation orientieren wir uns in diesem Abschnitt an der Vorgehensweise, die vom Evaluationsprozess vorgegeben wird: Zunächst werden in Abschnitt 10.1.1 mit Hilfe des Interaktiven Kooperationsturniers viel versprechende Gegenstrategien gefunden, deren Einsatz von manipulationswilligen Benutzern zu erwarten ist. Anschließend werden die gefundenen Gegenstrategien in Abschnitt 10.1.2 unter Zuhilfenahme des Simulativen Kooperationsturniers untereinander verglichen und bewertet. Ziel ist es herauszufinden, unter welchen Rahmenbedingungen welche Gegenstrategie am erfolgreichsten ist. Durch die Berücksichtigung der jeweils besten Gegenstrategie in der Evaluation des Gesamtsystems wird eine realitätsnahe und aussagekräftige Simulation des Informationssystems ermöglicht.

10.1.1 Findung von Gegenstrategien

In diesem Abschnitt wird das Ergebnis der Anwendung des Interaktiven Kooperationsturniers besprochen. Hierzu geben wir zunächst einen Überblick der Versuchsdurchführung. Anschließend werden die gefundenen Gegenstrategien vorgestellt. Dabei unterscheiden wir zwischen einfachen und komplexen Gegenstrategien.

Versuchsdurchführung. Mit Hilfe des Interaktiven Kooperationsturniers wurden etwa 40 Turniere ausgetragen. Diese Turniere fanden in dem Zeitraum eines Jahres je nach Bedarf in einer Frequenz zwischen einmal pro Monat und mehrmals wöchentlich statt. Jedes einzelne Turnier wurde bis auf wenigen Ausnahmen in einer Sitzung von ungefähr zwei Stunden durchgeführt. An einem Turnier waren durchschnittlich 30 Einheiten beteiligt. Trotz dieser relativ hohen Zahl waren die Versuchspersonen durch den Einsatz der Avatare und des Verhaltensprotokolls in der Lage, die einzelnen Einheiten wiederzuerkennen und sich an ihr vergangenes Verhalten zu erinnern. Insgesamt haben über 30 verschiedene Versuchspersonen am Interaktiven Kooperationsturnier teilgenommen und darin spielerisch Gegenstrategien entwickelt. Zehn von ihnen haben in mehreren Turnieren mitgewirkt.

Bei jedem der Turniere wurden die Rahmenbedingungen variiert, um verschiedenartig ausgerichtete Gegenstrategien zu erhalten. So lag zum Beispiel die Rundenzahl im Bereich zwischen 5 und 20. Bei einer kleinen Rundenzahl entwickelten sich kurzfristig angelegte Gegenstrategien. Im Gegensatz dazu waren die gefundenen Gegenstrategien bei einer hohen Rundenzahl eher komplexerer Natur und langfristig angelegt. Auch das Nutzen-/Kostenverhältnis, die Zahl potentieller Transaktionspartner und der Anteil normativer Einheiten hatte Einfluss auf die entwickelten Gegenstrategien.

In den nachfolgenden Paragraphen werden die gefundenen Gegenstrategien vorgestellt. Sie lassen sich grob in zwei Gruppen einteilen: Die *einfachen* Gegenstrategien wurden von den meisten Versuchspersonen bei ihren ersten Teilnahmen an Turnieren entwickelt. Sie zeichnen sich dadurch aus, dass sie nahe liegende Kriterien heranziehen, um zu bestimmen, ob Betrugsverhalten sinnvoll ist. Hingegen sind die Gegenstrategien aus der zweiten Gruppe in der Hinsicht *komplex*, als sie sich erst nach einer Vielzahl von Turnieren herauskristallisiert haben. Bei ihrer Entwicklung haben die Versuchspersonen von ihrer Turniererfahrung gezehrt. Diese komplexen Gegenstrategien haben gemein, dass sie auf die initiale Bildung von Vertrauen einen großen Wert legen, um zu ausgewählten Zeitpunkten effektiver betrügen zu können.

Einfache Gegenstrategien. Die einfachste aller gefundenen Gegenstrategien ist AllDefector. Sie entspricht der Strategie *allD* der evolutionären Spieltheorie aus Abschnitt 4.1.3. Wie der Name andeutet, führt die Verfolgung dieser Gegenstrategie dazu, dass eine Einheit in allen ihrer Transaktionen betrügt. Bemerkenswert bei dieser Gegenstrategie ist das Folgende: Nur sehr wenige Versuchspersonen haben diese Strategie verfolgt. Wenn sie das taten, dann nur bei ihrer ersten Teilnahme. Es wurde den Versuchspersonen nämlich deutlich, dass sich nur durch zum Teil normativen Verhalten hinreichend viele Gelegenheiten zum Betrug ergeben. In dieser Hinsicht erscheint AllDefector als Gegenstrategie nicht sehr viel versprechend[1]. Sie wird daher durch das Aufstellen bestimmter Betrugskriterien von den

[1]Dennoch wird bei fast allen bestehenden Arbeiten zur verteilten Vertrauensbildung, die in Abschnitt 2.4 besprochen sind, die Gegenstrategie AllDefector als einzige in die simulative Evaluation einbezogen. Aus der Vielzahl und Unterschiedlichkeit der gefundenen und in diesem Abschnitt besprochenen Gegenstrategien wird somit deutlich, dass die Evaluationsmethodik dieser Arbeit einen nicht zu vernachlässigen Beitrag darstellt.

nachfolgenden Gegenstrategien verfeinert.

Die Gegenstrategie ProbabilisticDefector nimmt den Gedanken auf, dass Einheiten durch zu häufigen Betrug zu schnell als strategisch identifiziert werden. Wird hingegen nur mit einer bestimmten Wahrscheinlichkeit p betrogen, so lässt sich vom jeweiligen Transaktionspartner beabsichtigter von unbeabsichtigtem Betrug schwerer unterscheiden. ProbabilisticDefector ist somit eine Gegenstrategie, die sich in p Transaktionen wie AllDefector und in $1 - p$ Transaktionen normativ verhält. Ein Nachteil dieser Gegenstrategie ist, dass ihr Betrugskriterium zu einfach gestaltet ist, als dass die Eigenheiten der Rahmenbedingungen einer Transaktion ausgenutzt werden könnten.

Diese Einsicht findet im CleverDefector Eingang: Diese Gegenstrategie orientiert sich an der Rolle, die eine Einheit in einer Transaktion spielt. Nur wenn sie sich in der sicheren Position findet, betrügt sie. Hingegen verzichtet sie in der Risiko-Position auf Betrugsverhalten. Dies ist in der Hinsicht durchdacht, als der eigene Betrug in der Risiko-Position gemäß dem Sechs-Wege Transaktionsprotokoll zu einem Abbruch der Transaktion vor der Aktionsausführung des Gegenübers führt. Somit ergibt sich nur in der sicheren Position ein Betrugsvorteil. Daher sind die nachfolgend vorgestellten Gegenstrategien allesamt Verfeinerungen von CleverDefector. Sie geben Kriterien an, unter denen selbst in der sicheren Position kein Betrug beabsichtigt ist.

ContextDefector ist eine Gegenstrategie, die als Betrugskriterium den Transaktionskontext heranzieht. Je höher dieser ist, desto eher wird betrogen, da dann der Betrugsvorteil besonders groß ist. Auf den ersten Blick erscheint es sinnvoll, dass das Betrugskriterium durch Angabe eines Schwellwertes definiert wird. Demnach kommt es genau dann zum Betrug, wenn die Höhe des Transaktionskontextes diesen Schwellwert überschreitet. In Rücksprache mit den Versuchspersonen hat sich jedoch gezeigt, dass ContextDefector anders zu formulieren ist: Für jeden Transaktionskontext gibt es eine gewisse Wahrscheinlichkeit des Betruges. Diese steigt proportional mit dem Transaktionskontext. In Analogie zum ProbabilisticDefector hat dies zum Vorteil, dass beabsichtigtes Betrugsverhalten schlechter von unbeabsichtigtem unterschieden werden kann. Somit wird eine Einheit, die diese Gegenstrategie verfolgt, nicht so schnell als strategisch identifiziert. ContextDefector erscheint als sehr viel versprechend, da nur dann betrogen wird, wenn der Betrugsvorteil besonders hoch ist. So gesehen ist es kein Zufall, dass diese Gegenstrategie im Interaktiven Kooperationsturnier von Versuchspersonen entwickelt wurde. Andererseits wurde die Möglichkeit zu kontextabhängigem Fehlverhalten im Entwurf der Glaubensbildung aus Abschnitt 6.3 bereits berücksichtigt. Die Bewertung der Gegenstrategien in Abschnitt 10.1.2 wird zeigen, ob ContextDefector als Gegenstrategie dennoch erfolgreich ist.

Bei der Entwicklung der Gegenstrategie DiscriminatoryDefector wurden die Versuchspersonen von einer anderen Einsicht geleitet: Aufgrund der Möglichkeit, den eigenen Transaktionspartner unter mehr als einer Einheit auswählen zu können, reicht es aus, nur mit einem Teil der Einheiten Vertrauen aufzubauen. Die anderen Einheiten können betrogen werden, ohne dass die Betrugskosten allzu hoch sind. Die Gegenstrategie DiscriminatoryDefector teilt die Einheiten daher in zwei Gruppen ein. Den Einheiten der ersten Gruppe gegenüber wird stets normatives Verhalten gezeigt. Hingegen werden die Einheiten der zweiten Gruppe immer betrogen. Bei der Gegenstrategie lässt sich einstellen, welcher Anteil p der Einheiten in die erste Gruppe fällt. Wie lässt sich die Einteilung der Einheiten dann vornehmen? Da hierbei die Versuchspersonen vom Zufall geleitet wurden, wird bei der Formulierung von DiscriminatoryDefector folgendermaßen vorgegangen: Aus den Identifikatoren der Einheiten wird je ein Hash-Wert berechnet. Außerdem wird in der Wertemenge der Hashfunktion ein Teil der relativen Größe p zufällig bestimmt. Fällt der Hash-Wert des Identifikators einer Einheit in diesen Bereich, so gehört die Einheit der ersten Gruppe an. Mit dieser Vorschrift ist gewährleistet, dass für die Verfolgungen

dieser Gegenstrategie kein Wissen um die Gesamtzahl der Einheiten vorhanden sein muss, um einen bestimmten Teil der Einheiten zu diskriminieren.

Komplexe Gegenstrategien. Die im Folgenden vorgestellten Gegenstrategien haben gemein, dass sie Betrugskriterien aufstellen, deren Sinn nicht unmittelbar ersichtlich ist. Dennoch haben diese Gegenstrategien im Allgemeinen weitaus besser als die einfachen Gegenstrategien abgeschlossen. Sie legen einen großen Wert auf die initiale Bildung von Vertrauen, um zu ausgewählten Zeitpunkten effektiver betrügen zu können. Wie diese Bildung von Vertrauen im Einzelnen aussieht, wird im Folgenden für jede komplexe Gegenstrategie besprochen.

DistrustDistributor ist eine Gegenstrategie, bei der Fehlverhalten an die eigenen Bürgschaftsbeziehungen ausrichtet ist. Damit versucht diese Gegenstrategie, den im Systementwurf vorgesehenen Einsatz sozialer Beweismittel für ihre Zwecke auszunutzen. Dabei sieht die Gegenstrategie folgendes Vorgehen vor: Die Einheit verhält sich normativ, bis sie eine bestimmte Anzahl von Bürgen besitzt. Anschließend betrügt sie so lange, bis diese Anzahl von Bürgen wieder unterschritten wird. Ihr Verhalten ist also in einem Zyklus zwischen rein normativem und rein betrügendem Verhalten organisiert, wobei die Phasen des Zyklus durch die Zahl der eigenen Bürgen gesteuert werden. Warum macht eine solche Gegenstrategie Sinn? Gemäß der Revisionsvorschriften zum sozialen Typglauben aus Abschnitt 8.3.2 wird bei negativen Transaktionserfahrungen eine Einheit umso weniger abgewertet, je mehr Bürgen sie besitzt. Umgekehrt fällt eine Aufwertung des individuellen Typglaubens umso stärker aus, je weniger Bürgen es gibt. Durch die Verfolgung von DistrustDistributor orientiert sich eine Einheit also daran, Betrugsverhalten genau dann zu zeigen, wenn sie durch ihre Bürgen hinreichend abgesichert ist. Damit wird der Versicherungseffekt, der laut Abschnitt 8.4.2 von Bürgschaftsbeziehungen ausgelöst wird, gezielt ausgenutzt. Ein wichtiger Aspekt von DistrustDistributor ist, dass nur mit normativ erscheinenden Einheiten Bürgschaftsbeziehungen eingegangen werden. Genauer gesagt werden die Vorschriften zum Eingehen von Bürgschaftsbeziehungen vom Systementwurf unverändert übernommen. Damit bestätigt sich in dieser Gegenstrategie die Paradoxie strategischen Verhaltens, das in der Analyse strategischen Bürgschaftsverhaltens in Abschnitt 8.4.2 prognostiziert wurde. Ein Problem von DistrustDistributor stellt die Gewinnung der eigenen Bürgen dar. Nach der Phase des Betruges ist unter Umständen über eine lange Zeitdauer hinweg normatives Verhalten vonnöten, um wieder zu eigenen Bürgen zu gelangen. Daher lässt sich vermuten, dass sich eine Einheit durch die Verfolgung von DistrustDistributor langfristig gesehen in der überwältigenden Zahl von Transaktionen normativ verhält. Auch hierdurch bestätigt sich das Ergebnis der Analyse aus Abschnitt 8.4.2, dass der Einsatz von Bürgschaftsbeziehungen für strategische Einheiten den Zwang zu normativem Verhalten verstärkt.

Eine weitere Gegenstrategie, die der vorhergehenden ähnelt, ist SkimmingTrustDefector. In Übereinstimmung mit DistrustDistributor organisiert diese Gegenstrategie ihr Betrugskriterium in zwei alternierende Phasen. Allerdings gibt es auch einige wesentliche Unterschiede: Die Steuerung dieser Phasen basiert nicht auf den eigenen Bürgen sondern auf dem eigenen Typglauben. Genauer gesagt wird genau in solchen Transaktionen betrogen, in denen der Typglaube über den Transaktionspartner einen gewissen Schwellwert überschreitet. Anschaulich gesprochen werden also genau die Einheiten betrogen, die als hinreichend normativ erscheinen. Welchen Sinn hat dieses Verhalten? Ziel der Gegenstrategie ist es, trotz des eigenen Betrugsverhaltens von den anderen Einheiten als Partner in zukünftigen Transaktionen angenommen zu werden. Dies ist jedoch nur genau dann der Fall, wenn der Typglaube über einen selbst hoch genug ist. Betrugsverhalten kann sich eine Einheit demnach nur erlauben, wenn sie als so normativ angesehen wird, dass sie trotz einer Abwertung in Folge von Betrugsverhalten als hinreichend

normativ erscheint, um weiterhin als Transaktionspartner akzeptiert zu werden. Allerdings stellt es ein Problem dar, dass keine verlässlichen Informationen über den Glauben anderer Einheiten vorliegen. Damit lässt sich dieses Betrugskriterium nicht direkt anwenden. Um dieses Problem zu lösen, trifft SkimmingTrustDefector die Annahme, dass Typglaube symmetrisch ist, also $p_X(N_Y) = p_Y(N_X)$ gilt. Als Folge dieser Annahme werden diejenigen Einheiten betrogen, die als hinreichend normativ erscheinen. Jedoch ist diese Annahme falsch, da gemäß dem Systementwurf der Typglaube auch nicht annähernd symmetrisch ist[2]. Insofern erscheint diese Gegenstrategie nicht so viel versprechend wie DistrustDistributor.

Die Gegenstrategie NormativeRetaliator schlägt einen anderen Weg als die beiden bisher vorgestellten komplexen Gegenstrategien ein. Es kommt genau dann zum Betrug, wenn der Transaktionspartner als hinreichend strategisch erscheint. In Analogie zu SkimmingTrustDefector kommt hierfür ein Schwellwert zum Einsatz. Ist der Typglaube über den eigenen Transaktionspartner geringer als dieser Schwellwert, so wird er betrogen. Damit lässt sich diese Gegenstrategie mit der Strategie *TFT* aus der evolutionären Spieltheorie vergleichen. Die Idee hinter dieser Gegenstrategie ist, dass die Folgekosten von Betrug geringer sind, wenn der Transaktionspartner strategisch ist. Dass dem in der Tat so ist, haben wir bereits bei der Besprechung des ersten Entwurfsprinzips in Abschnitt 5.4.3 und der Analyse des Transaktions-, Empfehlungs- und Bürgschaftsverhaltens in Abschnitt 7.6.1, 7.6.2 und 8.4.2 gesehen. Die Befolgung der Gegenstrategie NormativeRetaliator ist jedoch insofern nicht ungefährlich, als auch eine strategisch erscheinende Einheit normativ sein und somit fälschlicherweise betrogen werden kann. Aus eben dieser Überlegung haben wir im Systementwurf darauf verzichtet, dass normative Einheiten betrügen, auch wenn ihr Transaktionspartner sehr strategisch erscheint. Dieser Verzicht wird dadurch ermöglicht, dass jede Einheit einer strategisch erscheinenden Einheit ihre Zustimmung zum Eingehen einer Transaktion versagen kann. Gemäß NormativeRetaliator wird hingegen keine Einheit als Transaktionspartner abgelehnt. Eben hierin liegt der Unterschied zwischen dem Systementwurf und dieser Gegenstrategie.

Im Laufe der ausgetragenen Turniere hat sich eine weitere Gegenstrategie herausgebildet, die TemptationDefector genannt wird. Sie kombiniert die Überlegungen aus drei bisher besprochenen Gegenstrategien. Bei ihnen handelt es sich um ContextDefector, SkimmingTrustDefector und NormativeRetaliator. Wenn das Betrugskriterium von einer dieser Gegenstrategien erfüllt ist, so fordert TemptationDefector Betrugsverhalten. Umgekehrt führt die Befolgung dieser Gegenstrategie nur dann zu normativem Verhalten, wenn keines der Betrugskriterien erfüllt ist. Besonders problematisch ist die Kombination von SkimmingTrustDefector und NormativeRetaliator: Nur Einheiten, deren Typ unklar erscheint, werden nicht betrogen. Es ist daher nicht auszuschließen, dass die Befolgung von TemptationDefector zu zu häufigem Betrugsverhalten führt. Ob dem so ist, muss die simulative Bewertung der Gegenstrategien zeigen, die im nachfolgenden Abschnitt angegangen wird.

10.1.2 Bewertung und Wahl der Gegenstrategien

Das Finden viel versprechender Gegenstrategien stellt den ersten Schritt dafür dar, die Richtungen der Manipulation zu antizipieren. In einem zweiten Schritt ist laut Evaluationsprozess

[2]Dies lässt sich am deutlichsten an den Revisionsvorschriften zu Bewertung von negativen Empfehlungen erkennen. Zwei an einem Konflikt beteiligte Einheiten Y und Z werden von der Einheit X abgewertet, ohne dass diese Einheit X von den anderen beiden Einheiten abgewertet würde. Somit ist $p_X(N_Y)$ in keinster Weise an $p_Y(N_X)$ gebunden.

für unterschiedliche Rahmenbedingungen zu ermitteln, welche der Gegenstrategien am besten abschneidet und daher zu erwarten ist.

Mit dieser Aufgabe beschäftigen wir uns in diesem Abschnitt. Hierzu wenden wir uns zunächst der Bewertung der Gegenstrategien zu. Ausgehend von den erhaltenen Ergebnissen halten wir fest, wie die Prinzipale strategischer Einheiten ihre Gegenstrategie wählen. Dabei gehen wir zunächst davon aus, dass die menschlichen Benutzer die Rahmenbedingungen, von denen diese Wahl abhängt, genau kennen. Anschließend wird das Resultat für den Fall erweitert, dass die Rahmenbedingungen nur unpräzise wahrgenommen werden können. Als Ergebnis erhalten wir die benötigte Abbildung zwischen Rahmenbedingungen und den darin zu erwartenden Gegenstrategien.

Bewertung der einzelnen Gegenstrategien. Um die gefundenen Gegenstrategien untereinander vergleichen zu können, ist eine Nachsimulation im Simulativen Kooperationsturnier nötig. In dieser muss sich zeigen, unter welchen Rahmenbedingungen welche Gegenstrategie am besten abschneidet. Die Güte einer Gegenstrategie können wir ermitteln, indem wir sie in einem Simulationslauf gegen normative Einheiten antreten lassen. Je höher die gemessenen Normativitätskosten sind, desto besser hat die Gegenstrategie abgeschnitten. Diese Simulationsläufe sind unter variierten Rahmenbedingungen durchzuführen, so dass wir die Abhängigkeit des Abschneidens von den Rahmenbedingungen bewerten können. Die in diesen Simulationsläufen erhaltenen Ergebnisse bilden die Grundlage für Aussagen darüber, für welche Gegenstrategien sich die Prinzipale strategischer Einheiten entscheiden. Die Bewertung der Gegenstrategien wurde in [Fäh05] durchgeführt. Im Folgenden stellen wir die wichtigsten der dort erhaltenen Ergebnisse vor.

Abbildung 10.1 zeigt vier Versuchsreihen, in denen jeweils die Zahl der Transaktionsgelegenheiten pro Einheit variiert wird. Die Versuchsreihen unterscheiden sich in je einer weiteren Rahmenbedingung. Anhand der Abbildung lassen sich einige Einsichten aufzeigen:

- AllDefector und ProbabilisticDefector: Diese beiden Gegenstrategien schneiden schlecht ab. Dies liegt daran, dass die Betrugskriterien dieser Gegenstrategien zu oft erfüllt sind. Nur bei einem geringen Anteil normativer Einheiten kommen diese beiden Gegenstrategien in den Bereich der anderen Gegenstrategien. Insgesamt sind beide Gegenstrategien jedoch bei allen Rahmenbedingungen mindestens einer weiteren Gegenstrategie unterlegen.

- CleverDefector und DiscriminatoryDefector: Im kurzfristigen Bereich schneiden diese beiden Gegenstrategien besonders gut ab. Dabei ist zu erkennen, dass CleverDefector im sehr kurzfristigen Bereich (5 Transaktionsgelegenheiten) und DiscriminatoryDefector im weniger kurzfristigen Bereich (10 Transaktionsgelegenheiten) die jeweils besten Gegenstrategien darstellen. Gemäß der Besprechung aus Abschnitt 10.1.1 kommt dieses Ergebnis nicht überraschend: CleverDefector verlangt immer dann nach Betrug, wenn die Einheit in der sicheren Position ist und somit Betrug einen Vorteil mit sich bringt. DiscriminatoryDefector schwächt die Betrugskriterien weiter ab und ist damit weniger kurzfristig angelegt. Einzige Ausnahme zu dieser Betrachtung bilden Systeme mit einem höheren Anteil an strategischen Einheiten. In diesen zahlt sich das kooperative Verhalten, das von DiscriminatoryDefector einigen ausgewählten Einheiten gegenüber vorgeschrieben ist, nicht aus.

- DistrustDistributor und SkimmingTrustDefector: Diese Gegenstrategien gehen von einem sehr ähnlichen Ansatz aus. Sie sind langfristig angelegt und schreiben nur dann Betrugsverhalten vor, wenn sich die Einheit dies aufgrund ihrer Stellung erlauben

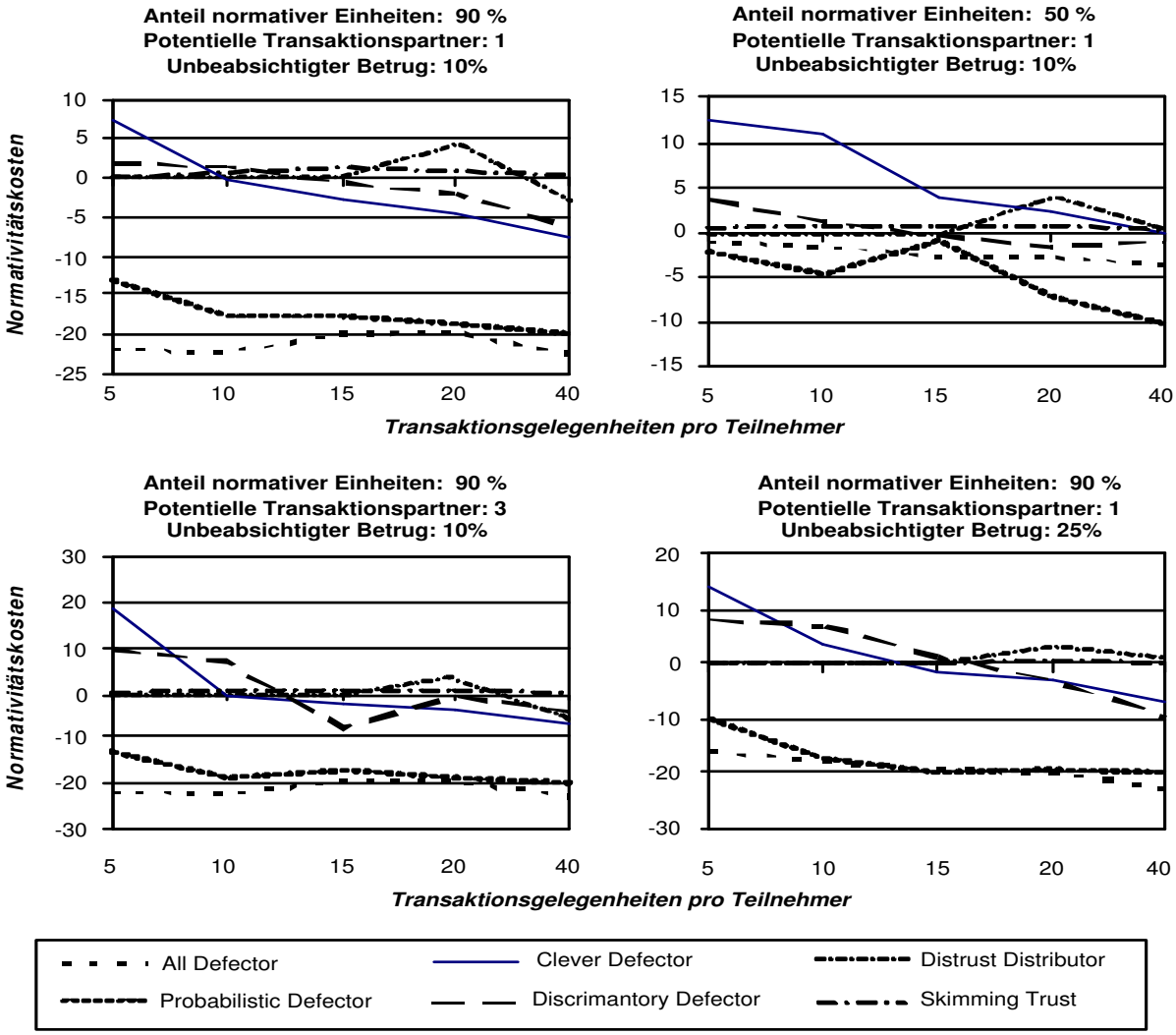

Abbildung 10.1: Ausgewählte Ergebnisse der Nachsimulation der Gegenstrategien

kann. Ein Vergleich dieser beiden Gegenstrategien ist daher besonders aufschlussreich. Gemäß DistrustDistributor werden zunächst Bürgschaftsbeziehungen aufgebaut und erst dann betrogen. Im kurzfristigen Bereich schreibt DistrustDistributor somit nur normatives Verhalten vor. Dies schlägt sich darin nieder, dass strategische Einheiten, die DistrustDistributor befolgen, in kurz laufenden Systemen (bis 15 Transaktionsgelegenheiten) genauso wie normative Einheiten abschneiden. Das Betrugsverhalten tritt massiv zu einem späteren Zeitpunkt (um 20 Transaktionsgelegenheiten) mit dem Ergebnis ein, dass die DistrustDistributor befolgenden Einheiten das in sie gesetzte Vertrauen verspielt haben und in der Folge nur schwerlich wieder Anschluss finden. Bei SkimmingTrustDefector sind die Phasen kooperativen und betrügenden Verhaltens kürzer. Daher besitzen die strategischen Einheiten, die SkimmingTrustDefector befolgen, stets einen kleinen Vorteil gegenüber normativen Einheiten. Besonders im langfristigen Bereich (ab 40 Transaktionsgelegenheiten) ergibt sich dadurch ein besseres Abschneiden als DistrustDistributor. Insgesamt lässt sich festhalten, dass beide Gegenstrategien längerfristig besser als die an-

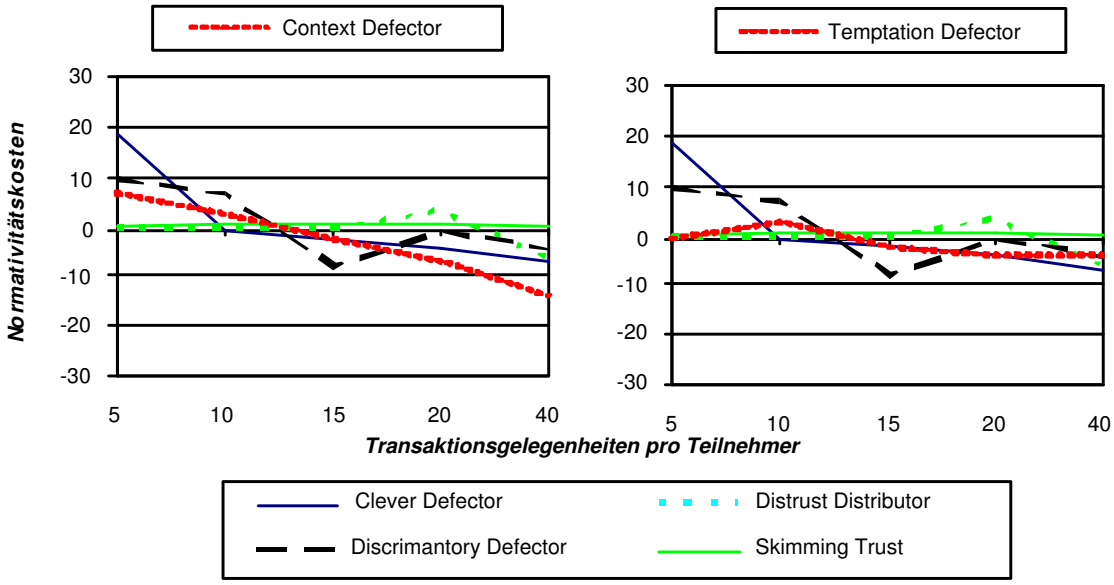

Abbildung 10.2: Einordnung des ContextDefector und TemptationDefector

deren Gegenstrategien abschneiden.

Weitere Versuchsreihen haben gezeigt, dass ContextDefector und TemptationDefector unter allen Rahmenbedingungen von jeweils einer anderen Gegenstrategie dominiert werden. Abbildung 10.2 stellt zwei Versuchsreihen dar, die dies zeigen. In ihnen werden die bisher besprochenen Gegenstrategien jeweils mit ContextDefector oder TemptationDefector verglichen:

- ContextDefector: Es ergibt sich ein auf den ersten Blick überraschender Verlauf. An sich ist zu erwarten, dass diese Gegenstrategie aufgrund des gemäßigten Betrugsverhaltens ähnlich wie DiscriminatoryDefector im mittelfristigen Bereich ihre Stärken besitzt. In der Tat schneidet dort ContextDefector im Vergleich zu anderen Gegenstrategien noch am Besten ab. Für alle Rahmenbedingungen gibt es jedoch zumindest eine andere Gegenstrategie, die besser als ContextDefector ist. Die Erklärung hierfür haben wir bereits in Abschnitt 10.1.1 gegeben: Die Möglichkeit zu kontextabhängigem Fehlverhalten wird im Systementwurf berücksichtigt. Dadurch werden die Vorteile aus kontextabhängigem Betrugsverhalten nivelliert. Anschaulich gesprochen betrügt eine strategische Einheit gemäß ContextDefector genau dann, wenn sie dafür besonders stark abgewertet wird (hoher Transaktionskontext). Umgekehrt verhält sie sich dann kooperativ, wenn durch normatives Verhalten nur eine kleine Aufwertung zu erwarten ist (geringer Transaktionskontext). Dies ist die Ursache für das schlechte Abschneiden.

- TemptationDefector: Wie erwartet wird aufgrund der Kombination unterschiedlicher Betrugskriterien zu häufig betrogen, als dass diese Gegenstrategie im langfristigen Bereich erfolgreich wäre. Darüber hinaus wird TemptationDefector auch im kurz- und mittelfristigen Bereich von den einfachen Gegenstrategien dominiert, da Betrugsverhalten nicht konsequent genug durchgehalten wird. Als Ergebnis erhalten wir, dass auch TemptationDefector für keine der Rahmenbedingungen die best mögliche Gegenstrategie darstellt.

Die Gegenstrategie NormativeRetaliator nimmt eine Sonderstellung ein, da sie gezielt das Betrugsverhalten anderer Gegenstrategien ausnutzt. Daher ist ein Vergleich normativen Einheiten gegenüber nicht sinnvoll. Vielmehr ist die Güte von NormativeRetaliator in Abhängigkeit der anderen beteiligten Gegenstrategien zu bewerten. Auf diese Aufgabe kommen wir im nachfolgenden Paragraphen zurück.

Wahl bei perfektem Wissen um die Rahmenbedingungen. Im Folgenden gehen wir davon aus, dass manipulationswillige Benutzer die Rahmenbedingungen des Informationssystems genau kennen. Insbesondere sind sie sich über den Anteil n der normativen Einheiten, die Zahl o der durchschnittlichen Transaktionsgelegenheiten pro Einheit und die Zahl p der potentiellen Transaktionspartner im Bilde. Für welche Gegenstrategie würden sie sich dann entscheiden? Die Antwort liegt in den zuvor besprochenen Versuchsreihen. Wir müssen davon ausgehen, dass sich die Benutzer für diejenige Gegenstrategie entscheiden, die für die jeweiligen Rahmenbedingungen am besten abschneidet[3].

In [Fäh05] werden ausgehend von dieser Vorüberlegung für jede Gegenstrategie diejenigen Rahmenbedingungen identifiziert, in denen sie die bestmögliche Wahl eines manipulationswilligen Benutzers ist. Dabei werden zwei Bereiche der Rahmenbedingungen unterschieden. Der *kurzfristige* Bereich erstreckt sich über alle Rahmenbedingungen, bei denen gilt:

$$(o < 11 \wedge n > 90\%) \vee (o < 17 \wedge 50\% < n \leq 90\%) \vee (o < 20 \wedge n \leq 10\%) \tag{10.1}$$

Hieraus wird offensichtlich, dass die Definition des kurzfristigen Bereichs nicht nur von der Zahl der Transaktionsgelegenheiten sondern auch vom Anteil der normativen Einheiten abhängt. Der Bereich ist umso größer, je kleiner der Anteil der normativen Einheiten ist. Die Rahmenbedingungen, für die diese Bedingung nicht erfüllt ist, nennen wir den *langfristigen* Bereich. Für jeden der beiden Bereiche gestaltet sich die Wahl der Gegenstrategie wie folgt:

- *Kurzfristiger Bereich:* Hierbei kommen CleverDefector und DiscriminatoryDefector in Frage. Aufgrund des Verlaufs in Abbildung 10.1 erhalten wir, dass DiscriminatoryDefector für diejenigen Rahmenbedingungen die beste Wahl ist, in denen $(p \geq 3 \wedge o > 7)$ oder $(n \geq 90\% \wedge o > 9)$ gilt. Ansonsten ist CleverDefector erfolgreicher.

- *Langfristiger Bereich:* In den Versuchsreihen schneidet SkimmingTrustDefector längerfristig besser als DistrustDistributor ab. Dies führt zur Wahl von SkimmingTrustDefector in solchen Rahmenbedingungen, in denen $(o \leq 16 \wedge n \geq 90\%)$, $(o \geq 40 \wedge n > 50\%)$ oder $(o > 65)$ gilt. Ansonsten schneidet DistrustDistributor besser ab.

Durch diese Beschreibung der Wahlkriterien manipulationswilliger Benutzer erhalten wir eine Abbildung Γ von den Rahmenbedingungen R zu der jeweils besten Gegenstrategie g. Auf diese Formalisierung der Abbildung als $\Gamma : R \rightarrow g$ kommen wir im nachfolgenden Paragraphen zurück.

Die Untersuchung von NormativeRetaliator in [Fäh05] zeigt, dass diese Gegenstrategie nur im langfristigen Bereich sinnvoll ist. Im kurzfristigen Bereich lohnt es sich nämlich nicht,

[3]Damit gehen wir davon aus, dass die manipulationswilligen Benutzer vor der Wahl der Gegenstrategie in der Lage sind, die Gegenstrategien etwa mit Hilfe eines Simulationswerkzeug selbst zu bewerten. Offensichtlich überschätzen wir damit die Fähigkeiten der manipulationswilligen Benutzer. Insofern gehen wir aus der Sicht des Systementwerfers von eine pessimistischen Annahme aus. Diese ist notwendig, damit die Ergebnisse der simulativen Evaluation auch dann gelten, wenn manipulationswillige Benutzer wider Erwarten über solche Fähigkeiten verfügen.

vom Betrugsverhalten normativen Einheiten gegenüber abzusehen. Im langfristigen Bereich ist diese Gegenstrategie hingegen durchaus erfolgreich. Dabei ist die Voraussetzung, dass die Gegenstrategie, die gemäß der Abbildung Γ gewählt wird, weiterhin von den meisten strategischen Einheiten verwendet wird. Der Anteil der strategischen Einheiten, die NormativeRetaliator befolgen, muss hingegen eher klein sein. Sonst gibt es für diese Einheiten zu wenige Gelegenheiten zu Betrugsverhalten. Diese Erkenntnis setzen wir wie folgt um: Liegen die Rahmenbedingungen R im langfristigen Bereich, so wird die Gegenstrategie $\Gamma(R)$ von 70% der strategischen Einheiten verfolgt. Die restlichen 30% strategischen Einheiten befolgen NormativeRetaliator. In diesem Verhältnis von 70 : 30 ist gemäß [Fäh05] ein Gleichgewicht zwischen der Gegenstrategie $\Gamma(R)$ und NormativeRetaliator antreffen: Ist es höher, so haben manipulationswillige Benutzer einen Vorteil darin, NormativeRetaliator zu wählen. Andererseits ist bei einem kleineren Verhältnis die Gegenstrategie $\Gamma(R)$ von Vorteil.

Wahl bei imperfektem Wissen um die Rahmenbedingungen. Bisher sind wir davon ausgegangen, dass manipulationswillige Benutzer die Rahmenbedingungen des Informationssystems genau kennen. Dies führte uns zur Formulierung einer Abbildung $\Gamma : R \rightarrow g$, die die Wahl manipulationswilliger Benutzer antizipiert. Im Folgenden befassen wir uns mit einer Verallgemeinerung dieser Abbildung. Dabei berücksichtigen wir, dass die Benutzer realistischerweise die Rahmenbedingungen a priori nicht exakt einschätzen können.

Die Grundidee ist die Folgende: Die manipulationswilligen Benutzer unterscheiden sich darin, wie sie die Rahmenbedingungen einschätzen. Wenn zum Beispiel die Zahl o der durchschnittlichen Transaktionsgelegenheiten 10 ist, gibt es Benutzer, die diese Zahl unter- beziehungsweise überschätzen. Die Einschätzungen liegen somit in einem Bereich $[l_o \cdot 10, r_o \cdot 10]$, wobei $l_o < 1$ und $r_o > 1$ gilt. Die beiden Faktoren l_o und r_o geben an, wie präzise die Benutzer die Rahmenbedingungen bezüglich der Zahl der Transaktionsgelegenheiten wahrnehmen. Sind beide Faktoren eins, so nehmen sie die Rahmenbedingungen exakt wahr und wir erhalten das Ergebnis aus dem vorigen Paragraphen. Allerdings müssen wir realistischerweise davon ausgehen, dass dem nicht der Fall ist. Dies gilt nicht nur für die Zahl der Transaktionsgelegenheiten, sondern zumindest auch für den Anteil n normativer Einheiten. Hierbei gibt es ebenfalls zwei Faktoren l_n und r_n, die die relative Bandbreite der Einschätzung bezüglich n festhalten.

Welche Auswirkung hat diese Grundidee auf die Wahl der manipulationswilligen Benutzer? Je nach ihrer individuellen Wahrnehmung schätzen sie die Rahmenbedingungen unterschiedlich ein. Unter Umständen werden deswegen verschiedene Gegenstrategien ausgewählt. Unterschätzt zum Beispiel ein Benutzer die Zahl der Transaktionsgelegenheiten, so entscheidet er sich möglicherweise zur kurzfristig angelegten Gegenstrategie CleverDefector. Ein anderer Benutzer, der die Zahl überschätzt, wird sich eher für die längerfristigen Gegenstrategien SkimmingTrustDefector oder DistrustDistributor entscheiden. Somit kommt es zu einer *Heterogenisierung* der zu erwartenden Gegenstrategien. Wir müssen also die Abbildung Γ zu einer Abbildung Γ_w erweitern, die von unterschiedlichen Wahrnehmungen seitens der manipulationswilliger Benutzer ausgeht. Das Ergebnis dieser Abbildung kann nicht eine einzelne Gegenstrategie sein. stattdessen muss die Abbildung angeben, wie hoch der Anteil manipulationswilliger Benutzer ist, die eine bestimmte Gegenstrategie g wählen. Somit ist Γ_w eine Funktion der Rahmenbedingungen R (angegeben durch n_0 und o_0) und der Faktoren der Wahrnehmung l_n, r_n, l_o und r_o. Die Abbildung ist wie

folgt definiert (δ bezeichnet die Dirac-Funktion[4]):

$$\Gamma_w(R,g) = \frac{1}{[(r_n - l_n)\cdot n_0]\cdot[(r_o - l_o)\cdot o_0]} \cdot \int_{l_n\cdot n_0}^{r_n\cdot n_0} \int_{l_o\cdot o_0}^{r_o\cdot o_0} \delta(g,\Gamma(n,o))\ do\ dn \tag{10.2}$$

Diese Definition erklärt sich folgendermaßen: Berechnet werden soll der Anteil der Benutzer, die sich zur Gegenstrategie g entscheiden. Hierfür ist zunächst der Wahrnehmungsbereich zu bestimmen. Dieser ergibt sich aufgrund der Faktoren der Wahrnehmung durch $[l_n\cdot n_0, r_n\cdot n_0]\times[l_o\cdot o_0, r_o\cdot o_0]$. Die Größe dieses Bereichs wird im Nenner des Vorfaktors berechnet. Anschließend wird die Größe desjenigen Teilbereichs bestimmt, in dem die Wahl auf die Gegenstrategie g fällt. Somit gibt $\Gamma_w(R,g)$ den Anteil der Ausprägungen der Wahrnehmungen an, bei die Gegenstrategie g gewählt wird. Dieser lässt sich als der Anteil derjenigen manipulationswilligen Benutzer interpretieren, die sich für die Gegenstrategie g entscheiden.

Zur Illustration geben wir ein Beispiel zur Berechnung von $\Gamma_w(R,g)$ an. Der Übersichtlichkeit wegen nehmen wir dabei an, dass nur die Zahl der Transaktionsgelegenheiten unterschiedlich eingeschätzt wird. Die eigentlichen Rahmenbedingungen seien durch $(n_0, o_0) = (70\%, 16)$ gegeben. Zudem seien die Faktoren der unpräzisen Wahrnehmung $l_o = 0,5$ und $r_o = 2$, so dass die Wahrnehmung von o im Intervall [8,32] liegt. Somit vereinfacht sich die Berechnung von Γ_w zu folgender Formel:

$$\frac{1}{32-8} \cdot \int_8^{32} \delta(g,\Gamma(n_0,o))\ do \tag{10.3}$$

Bei einem Anteil von normativen Einheiten von 70% wird laut der Abbildung Γ der Bereich kurzfristiger Gegenstrategien bei $o = 17$ von dem der langfristigen Gegenstrategien getrennt. Wir erhalten je nach Wahrnehmung drei mögliche Gegenstrategien. Diese lauten mitsamt ihres Anteils am Intervall [8,32] wie folgt:

- $o < 9$: CleverDefector. Für diese Gegenstrategie g_1 ist $\Gamma_w(R,g_1) = \frac{9-8}{32-8} = \frac{1}{24} = 4,17\%$.
- $9 \leq o < 17$: DiscriminatoryDefector. Hierbei ist $\Gamma_w(R,g_2) = \frac{17-9}{32-8} = \frac{1}{3} = 33,33\%$.
- $17 \leq o$: DistrustDistributor. Wir erhalten $\Gamma_w(R,g_3) = \frac{32-17}{32-8} = \frac{15}{24} = 62,50\%$.

SkimmingTrustDefector erhält keinen Anteil, da diese Gegenstrategie gemäß der Rahmenbedingung nur bei einer Wahrnehmung von $o > 40$ gewählt werden würde. Dies ist aber nicht der Fall, da die Wahrnehmung bezüglich o maximal 32 beträgt.

Aus dem Beispiel wird deutlich, dass die Abbildung $\Gamma_w(R,g)$ sinnvoll definiert ist: Benutzer wählen nur CleverDefector, wenn sie die Zahl der Transaktionsgelegenheiten stark unterschätzen. Entsprechend gering fällt der Anteil dieser Strategie aus. Der Anteil der Gegenstrategie DistrustDistributor ist hoch, da eine leichte Überschätzung der Zahl der Transaktionsgelegenheiten für ihre Wahl ausreicht. Wären wir hingegen von der Annahme ausgegangen, dass die Rahmenbedingungen perfekt wahrgenommen werden, so wäre DiscriminatoryDefector als die einzige Gegenstrategie gewählt worden. Dies ist insofern fraglich, als bei einer leicht erhöhten Zahl der Transaktionsgelegenheiten ($o = 17$) alle Benutzer DistrustDistributor wählen würden. Durch die Berücksichtigung unpräziser Wahrnehmung erhalten wir also einen stetigen Übergang zwischen den Anteilen unterschiedlicher Gegenstrategien. Damit wird letztendlich die Realitätsnähe der simulativen Evaluation erhöht.

[4] Die Dirac-Funktion δ ist eine Abbildung $\mathcal{A}^2 \to \{0,1\}$, die für alle (a,a) mit $a \in \mathcal{A}$ eins und sonst null zurückgibt.

10.2 Evaluation des Gesamtsystems

Die Antizipation der Manipulation bildet die Grundlage dafür, dass das Gesamtsystem auf realistische Weise simuliert werden kann. In diesem Abschnitt wenden wir uns in einem zweiten Schritt dem eigentlichen Ziel der Evaluation zu: Durch die Simulation des Gesamtsystems im Simulativen Kooperationsturnier zeigen wir, unter welchen Rahmenbedingungen unsere These von der Existenzfähigkeit des Informationssystems validiert werden kann. Eine Übersicht der zu diesem Zweck durchgeführten Versuche gibt Abschnitt 10.2.1.

Die Darstellung und Interpretation der Versuche ist nach unterschiedlichen Gesichtspunkten gegliedert: Abschnitt 10.2.2 gibt Aufschluss darüber, unter welchen Rahmenbedingungen das Informationssystem existenzfähig ist. Da die dafür eingesetzten Versuchsreihen den Anteil normativer Einheiten am Gesamtsystem variieren, lassen sich dabei Rückschlüsse auf die Populationsstruktur existenzfähiger Informationssysteme ziehen. Die Abhängigkeit einiger ausgewählter Dimensionen der Rahmenbedingungen und ihre Auswirkungen auf die Existenzfähigkeit des Informationssystems werden in Abschnitt 10.2.3 untersucht. In Abschnitt 10.2.4 gehen wir der Frage nach, welchen Einfluss die menschlichen Eigenschaften und Präferenzen auf die Existenzfähigkeit des Informationssystems ausüben. Abschließend werden die Ergebnisse der Versuche in Abschnitt 10.2.5 in Zusammenhang gebracht. Das Fazit, das wir dabei ziehen, gibt Auskunft darüber, inwiefern unsere These von der Existenzfähigkeit des Informationssystems validiert werden kann.

10.2.1 Versuchsübersicht

Die Evaluationsmethodik gibt gemäß Abschnitt 9.1.2 die Ausrichtung der Gesamtevaluation vor. Ausgangspunkt ist das dortige Kriterium, mit dem die Existenzfähigkeit des Informationssystems beurteilt werden kann. Basierend darauf wird das Informationssystem unter unterschiedlichen Rahmenbedingungen simuliert. Die gezielte Variation einzelner Aspekte der Rahmenbedingungen ermöglicht Aussagen darüber, wie die Existenzfähigkeit des Informationssystems von den Eigenheiten der Rahmenbedingungen abhängt. Eine Übersicht der dazu durchgeführten Versuchsreihen muss also auf die jeweilige Einstellung der Rahmenbedingungen eingehen.

Mit dieser Aufgabe befassen wir uns in diesem Abschnitt. Zunächst gehen wir auf die Standardeinstellung ein, die den Ausgang für gezielte Variationen der Rahmenbedingungen bildet. Anschließend untersuchen wir, in welche Richtungen zu evaluieren ist, um zu den Einsichten zu gelangen, die zur Verifikation unser These benötigt werden. Dabei wird eine Übersicht gegeben, für welche Aspekte der Rahmenbedingungen in den nachfolgenden Abschnitten eine Sensibilitätsanalyse durchgeführt wird.

Angaben zum Entwurfspunkt, der für die Gesamtevaluation gewählt wurde, werden in Abschnitt A.2.1 des Anhangs gemacht. Er setzt alle Mechanismen ein, die im Teil II dieser Arbeit entworfen worden sind. In diesem Abschnitt zielen wir also auf die ganzheitliche Evaluation des Entwurfes. Untersuchungen dazu, ob der Entwurf im Hinblick auf die von ihm eingeführten Mechanismen minimal ist, sind somit nicht Bestandteil der nachfolgenden Betrachtungen. Solche Untersuchungen könnten in zukünftigen Arbeiten durchgeführt werden. Weiterhin wird auf einen simulativen Vergleich mit einem Informationssystem ohne Vertrauensbildung verzichtet, da dessen mangelnde Existenzfähigkeit durch die analytischen Überlegungen der Abschnitte 1.2.1 und 5.3.2 hinreichend gezeigt worden ist.

Standardeinstellung der Rahmenbedingungen. Die Beschreibung der Simulationsumgebung aus Abschnitt 9.2.1 zeigt, dass die Rahmenbedingungen der Simulation von einigen Modell-

parametern abhängen. Diese belegen wir im Folgenden mit Standardwerten, die den nachfolgenden Simulationen (wenn nicht anders angegeben) zugrunde liegen.

Tabelle 10.1 gibt einen Überblick der Belegung der Modellparameter. Jedem von ihnen weist die Tabelle einen Kürzel zu, der für die Spezifikation der Versuchsreihen in den weiteren Abschnitten Verwendung findet wird. Im Folgenden gehen wir auf die Belegungen im Einzelnen ein. Dabei orientieren wir uns an die Überlegungen, die wir im Hinblick auf die Simulationsumgebung in Abschnitt 9.2.1 angestellt haben:

- *Kooperations- und Kommunikationsmodell:* Das durchschnittliche Nutzen-/Kostenverhältnisses von Aktionen sowie die durchschnittliche Zahl der Transaktionsgelegenheiten stellen konservative Schätzungen dar. Dadurch tragen wir Rechnung, dass nicht alle Studenten so aktiv und häufig wie zum Beispiel Anna das Informationssystem für ihre Zwecke verwenden müssen. Auch bei der Festlegung des Anteils der potentiellen Transaktionspartner berücksichtigen wir, dass zwar Informationen wie der Mensaplan weit verfügbar sind, andere Informationen wie zum Beispiel Vorlesungsmitschriebe nur von wenigen Einheiten gehalten werden. Gemäß Abschnitt 9.2.1 bezieht sich dieser Anteil auch auf diejenigen Einheiten, die zwar die benötigte Information besitzen aber außer Reichweite sind. Er ist daher mit 10% sehr vorsichtig gewählt. Die Kosten für das Versenden einfacher Nachrichten sind im Verhältnis zu den durchschnittlichen Kosten für die Aktionsausführung angegeben. Wie wir in Abschnitt 9.2.1 bereits gesehen haben, liegen mehrere Größenordnungen zwischen diesen beiden Kostenarten. Somit stellt die Festlegung dieses Modellparameters ebenfalls eine konservative Schätzung dar. Dasselbe gilt für die Festlegung der Wahrscheinlichkeit davon, dass es während einer Aktionsausführung zu einem Kommunikationsabbruch und damit einem unbeabsichtigtem Fehler kommt. Wie in Abschnitt 9.2.1 gezeigt, ist selbst bei aufwändigen Aktionen ist diese Wahrscheinlichkeit sehr gering.

- *Modell der Einheiten:* Bei den meisten Versuchsreihen wird der Anteil der normativen Einheiten variiert, um gemäß Abschnitt 9.1.2 zu Schlüssen über die Güte des Entwurfs kommen zu können. Der in der Tabelle angegebene Wert für den Anteil normativer Einheiten bezieht sich daher lediglich auf diejenigen Versuchsreihen, in denen mehrere Dimensionen der Rahmenbedingungen gleichzeitig variiert werden und dabei der Anteil normativer Einheiten fest vorgegeben sein muss. Die Aussage solcher Versuchsreihen bezieht sich somit darauf, ob ein Informationssystem mit einem relativ großen Anteil von normativen Einheiten (70%) existenzfähig ist. Wie bereits in Abschnitt 9.2.1 besprochen, wird die Zahl der simulierten Einheiten auf 100 festgelegt. Die Skalierbarkeit des Informationssystems wird in speziell ausgerichteten Versuchsreihen untersucht, die diese Zahl variieren. Die Wahrnehmung der Rahmenbedingungen durch manipulationswillige Benutzer ist mit Hilfe der Faktoren aus Abschnitt 10.1.2 angegeben. Dabei gehen wir (zu unserem eigenen Nachteil) davon aus, dass diese Benutzer die Rahmenbedingungen relativ genau kennen und bei der Wahl der manipulierten Version diese Kenntnis auch ausnutzen.

Abschließend wird die Verteilung der Manipulationskosten in Abbildung 10.3 dargestellt. Entsprechend der Vorgaben aus Abschnitt 9.2.1 beruht sie auf einer äußerst vorsichtigen Schätzung. Laut Abschnitt 5.3.2 hält die Verteilung $F(m)$ fest, welcher Anteil der Benutzer mindestens die angegebenen Manipulationskosten m besitzt. Dabei werden die Manipulationskosten in durchschnittliche Kosten einer Aktionsausführung angegeben. Der Wert $F(10) = 50\%$ ist zum Beispiel so zu interpretieren, dass jeder zweite Benutzer sich zur Manipulation entscheidet, wenn er sich

Tabelle 10.1: Standardparametrisierung der Umgebung in den Versuchsreihen

Dimension	**Abkürzung**	**Ausprägung**
Durchschnittlicher Nutzen der Aktionsausführung im Vergleich zu ihren Kosten	b	3,5
Durchschnittliche Zahl von Transaktionsgelegenheiten pro Einheit	o	20
Durchschnittlicher Anteil potentieller Transaktionspartner am Gesamtsystem	p	10%
Kosten einer Nachricht im Vergleich zu den durchschnittlichen Kosten einer Aktionsausführung	k	0,02
Wahrscheinlichkeit eines unbeabsichtigten Fehlers bei der Aktionsausführung	u	5%
Anteil normativer Einheiten	n	70%
Zahl der Einheiten im System	s	100
Wahrnehmung der Rahmenbedingungen (Faktoren gemäß Abschnitt 10.1.2)	(l_n-r_n);(l_o-r_o)	(0,8-1,2);(0.5-2)
Verteilung der Manipulationskosten	F(m)	Abbildung 10.3

dadurch die Kosten für die Ausführung von 10 Aktionen (oder entsprechend mehr Aktionen wenn Betrugskosten anfallen) sparen kann. Die Verteilung ergibt sich aus der folgenden Definition:

$$F(m) = 1 - (\frac{1}{2})^{\frac{m}{10}} \tag{10.4}$$

Diese Verteilung der Manipulationskosten ist wie folgt zu interpretieren: **(1)** Wir nehmen an, dass die erste manipulierte Version erstellt und verwendet wird, sobald die Normativitätskosten auch nur unwesentlich größer als null sind. Das bedeutet für den Ersteller der manipulierten Version, dass ihm die Überwindung der technischen Hindernisse keine Kosten verursacht und er die rechtlichen Hindernisse ignoriert. Gemäß der Szenariobeschreibung fällt der Student Manuel in diese Gruppe von Benutzer. **(2)** Die Hälfte der Benutzer besitzt Manipulationskosten von unter 10. Anstatt die recht geringe Beanspruchung ihres Informationsgeräts für die Ausführung von zehn Aktionen hinzunehmen, ziehen es diese Benutzer somit vor, eine manipulierte Version zu erstellen oder zu übernehmen und die dabei anfallenden substantiellen Hindernisse zu überwinden. Aus der Szenariobeschreibung lässt sich der Student Bob dieser Gruppe von Benutzern zuordnen. **(3)** Die andere Hälfte der Benutzer besitzt Manipulationskosten über 10. Es gibt sogar einige Benutzer, deren Manipulationskosten so groß sind, dass sie lieber aus dem Informationssystem austreten, als sich den Hindernissen und Gefahren der Verwendung einer manipulierten Version auszusetzen. Beispiele für solche Benutzer sind Anna und Claude.

Diese Besprechung der Verteilung der Manipulationskosten zeigt, dass sie sich wie gefordert um eine vorsichtige Schätzung handelt. Auch diese Verteilung wird im Zuge der Gesamtevaluation in speziell ausgerichteten Versuchsreihen variiert.

Ausrichtung der Versuchsreihen. Mit Hilfe des Simulativen Kooperationsturniers lassen sich Sensibilitätsanalysen durchführen und wie in Abschnitt 9.2.2 graphisch darstellen. Der Übersichtlichkeit wegen empfiehlt es sich, bei jeder Sensibilitätsanalyse jeweils zwei Modellparameter

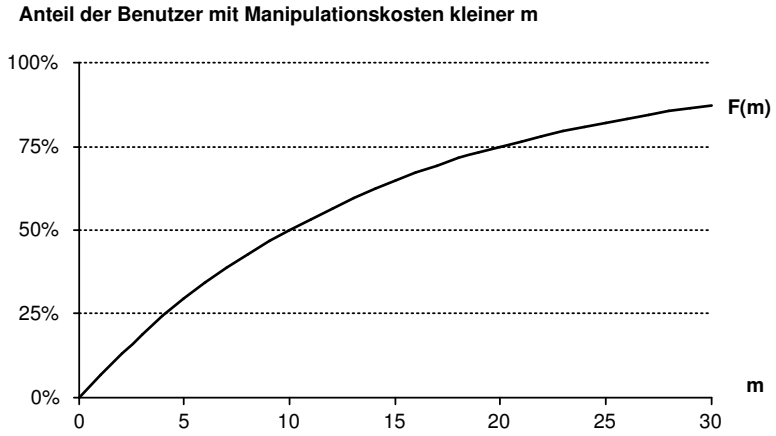

Abbildung 10.3: Verteilung der Manipulationskosten

Tabelle 10.2: Übersicht der Versuchsreihen zur Gesamtevaluation mit Abbildungsverzeichnis

Abbildung	Variation
10.4(a)	**n**, **o**
10.4(b)	**n**
10.5(a)	**n**, **b**
10.5(b)	**n**, **k**
10.6(a)	**n**, **p**
10.6(b)	**p**
10.7(a)	**n**, **s**
10.7(b)	**s**

Abbildung	Variation
10.8(a)	**n**, **u**
10.8(b)	**u**
10.9	**o**, **b**
10.10	**o**, **s**
10.11	**o**, **u**
10.12(a)	**n**, **o** (l_n=r_n=1, l_o=r_o=1)

Abbildung	Variation
10.12(b)	**n** (l_n=r_n=1, l_o=r_o=1)
10.14(a)	**n**, **o** ($F(m)$ gemäß Abbildung 10.3(a))
10.14(b)	**n**, **o** ($F(m)$ gemäß Abbildung 10.3(b))

der Rahmenbedingungen zu variieren. Zwar könnten mehrere solcher Parameter gleichzeitig untersucht werden. Zur Darstellung müsste das Ergebnis der Analyse jedoch wieder auf die Variation zweier Modellparameter zurückgeführt werden.

Eine Besonderheit stellen zwei Parameter des Modells der Einheiten dar. Es handelt sich bei ihnen um die Wahrnehmung der Rahmenbedingungen (l_n-r_n, l_o-r_o) und die Verteilung der Manipulationskosten $F(m)$. Diese Modellparameter sind keine einfachen Skalare, die im Sinne der Sensibilitätsanalyse variiert werden könnten. Bei den Versuchsreihen, die diese Modellparameter zu variieren sind, werden daher einige bestimmte ihrer Ausprägungen gezielt untersucht.

Die Versuchsreihen, die durchgeführt wurden, lassen sich in drei Gruppen unterteilen. Im Folgenden wird jede dieser Gruppen und die Ausrichtung ihrer Versuchsreihen vorgestellt:

- *Populationsstruktur (Abschnitt 10.2.2):* Diese Versuchsreihen haben gemein, dass sie eine Aussage über die Güte des Entwurfs treffen. Dies erfordert die Variation des Anteils der normativen Einheiten. Damit lassen sich Rückschlüsse auf die Populationsstruktur existenzfähiger Informationssysteme ziehen. Zusätzlich zum Anteil normativer Einheiten wird ein zweiter Modellparameter variiert.

- *Innere Abhängigkeit (Abschnitt 10.2.3):* Hierbei wird die Abhängigkeit einiger ausgewählter Modellparameter untereinander und ihre Auswirkung auf die Existenzfähigkeit des Informa-

tionssystems untersucht. Ein Beispiel hierfür ist die Versuchsreihe, in der die Abhängigkeit zwischen der Systemgröße und der Zahl der Transaktionsgelegenheiten herausgestellt wird.

- *Menschliche Eigenschaften und Präferenzen (Abschnitt 10.2.4):* In diesen Versuchsreihen vergleichen wir unterschiedliche Modellierungen der menschlichen Benutzer. Dies betrifft die Modellparameter zur menschlichen Wahrnehmung der Rahmenbedingungen und zur Verteilung der Manipulationskosten. Da diese Parameter nicht kontinuierlich variiert werden können, steht es uns offen, welche zwei weiteren Modellparameter in die Sensibilitätsanalyse eingehen.

Tabelle 10.2 gibt eine Übersicht aller Versuchsreihen der Gesamtevaluation. Hierbei ist für jede der nachfolgenden Abbildungen aufgetragen, welche Modellparameter variiert wurden. Die jeweils nicht variierten Parameter sind entsprechend der Standardeinstellung belegt.

Die Ausführung des Simulativen Kooperationsturniers für die einzelnen Versuchsreihen umfasste insgesamt über 65.000 Simulationsläufe. Die durchschnittliche Abarbeitungszeit eines Laufs betrug auf einem IBM ThinkPad T40p ungefähr drei Minuten. Für die Durchführung der hier vorgestellten Versuchsreihen waren somit fünf Monate ununterbrochener Simulation notwendig.

10.2.2 Populationsstruktur existenzfähiger Informationssysteme

In diesem Abschnitt werden die Versuchsreihen besprochen, die den Anteil normativer Einheiten variieren. Die Ergebnisse dieser Versuchsreihen erlauben Aussagen darüber, welche Populationsstruktur in einem existenzfähigen Informationssystem zu erwarten ist.

Die folgende Darstellung ist anhand der Modellparameter gegliedert, die zusätzlich zum Anteil der normativen Einheiten variiert werden. Die Modellparameter zu den menschlichen Eigenschaften und Präferenzen werden nicht in diesem sondern im nachfolgenden Abschnitt 10.2.4 variiert.

Durchschnittliche Zahl der Transaktionsgelegenheiten pro Einheit (o). Zunächst wenden wir uns dem Modellparameter zu, der darüber bestimmt, wie oft es zur Kooperation im Informationssystem kommt. Dazu wird die durchschnittliche Zahl der Transaktionsgelegenheiten pro Einheit o um ihre Standardeinstellung von 20 im Intervall $[5, 35]$ variiert. Das Ergebnis der Sensibilitätsanalyse ist in Abbildung 10.4(a) dargestellt. Im Folgenden wird dieses Ergebnis interpretiert.

Zunächst ist festzuhalten, dass für einen Anteil $n \leq 90\%$ das Informationssystem unabhängig von der Zahl der Transaktionsgelegenheiten existenzfähig ist. Dies wird dadurch ermöglicht, dass die Normativitätskosten in keinem Simulationslauf den Wert 2 übersteigen. Anschaulich gesprochen gibt es also unabhängig der Rahmenbedingungen keinen nennenswerten Vorteil für die Verwendung manipulierter Versionen der Systemsoftware. Weiterhin bedeuten niedrige Normativitätskosten, dass die inhärent manipulationsfreudigen Benutzer gegenüber diejenigen Benutzer, die die originale Systemsoftware benutzen, kaum einen Vorteil haben. Insofern ist es auch unter dem Gesichtspunkt der Fairness tragbar, dass nicht alle Einheiten normativ sind. Somit setzt sich die originale Systemsoftware im Gesamtsystem durch.

Wie kommt es zu den vereinzelten Punkten, in denen das Informationssystem nicht existenzfähig ist? Den Grund hierfür haben wir bereits in Abschnitt 9.1.2 besprochen: Bei der Verteilung der Manipulationskosten müssen wir davon ausgehen, dass es einige wenige Benutzer gibt, denen die Manipulation der originalen Systemsoftware kaum oder keine Kosten verursacht. Diese inhärent manipulationsfreudigen Benutzer entscheiden sich schon bei sehr geringen Normativitätskosten zur Manipulation. Es bleibt daher nur noch die Frage zu klären, warum die

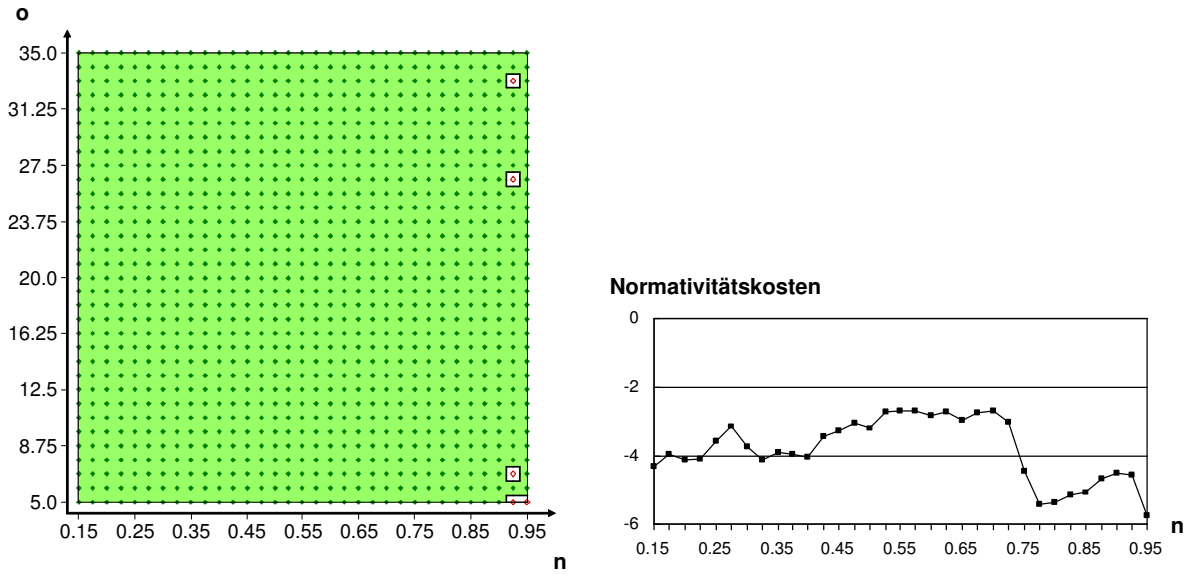

Abbildung 10.4: **(a)** Sensibilitätsanalyse bezüglich des Anteils der normativen Einheiten n und der durchschnittlichen Zahl der Transaktionsgelegenheiten pro Einheit o; **(b)** Abhängigkeit der Normativitätskosten von n

vereinzelten Punkte, in denen das Kriterium der Existenzfähigkeit nicht erfüllt ist, in zwei Bereichen ($o \approx 6$ und $25 < o < 35$) gehäuft auftreten. Die Antwort hierauf liegt in der Stärke der Gegenstrategien, die gemäß Abschnitt 10.1 je nach Eignung zur Simulation herangezogen werden. Bei wenigen Transaktionsgelegenheiten ist das Betrugsverhalten von CleverDefector am ehesten vorteilhaft. Der Einsatz dieser Gegenstrategie führt im Bereich $o \approx 6$ zu einigen Punkten, in denen das Kriterium verletzt ist. Analog dazu ist DistrustDistributor im Bereich $25 < o < 35$ erfolgreich und führt dort zu vereinzelten Verletzungen des Kriteriums. In diesem Bereich besitzen die strategischen Einheiten hinreichend viele Bürgen, um gemäß der Gegenstrategie in die Betrugsphase zu wechseln.

Eine andere Darstellung der Ergebnisse der Sensibilitätsanalyse wird in Abbildung 10.4(b) gewählt. Hierbei wird $o = 20$ festgehalten, um die Abhängigkeit der Normativitätskosten vom Anteil normativer Einheiten n zu illustrieren. Es ergibt sich, dass die Normativitätskosten dann am höchsten sind, wenn der Anteil normativer und strategischer Einheiten ungefähr gleich hoch ist ($n \approx 50\%$). In Systemen mit einem hohen Anteil an normativen Einheiten schneiden strategische Einheiten besonders schlecht ab. Dies bestätigt unsere Intuition, dass die soziale Kontrolle der Einheiten untereinander umso schärfer ist, je mehr Einheiten normativ sind.

Nutzen-/Kostenverhältnis der Kooperation (b und k). Eine wichtige Frage bei Evaluation des Gesamtsystems ist, wie das Verhältnis zwischen Nutzen und Kosten der Kooperation geartet sein muss, damit das Informationssystem existieren kann. Dieses Verhältnis wird durch zwei Modellparameter bestimmt. Bei ihnen handelt es sich um das durchschnittliche Nutzen-/Kostenverhältnis b der Aktionsausführung und dem durchschnittlichen Verhältnis k zwischen den Kosten einer Nachrichtenübermittlung und denen einer Aktionsausführung. Die Sensibilitätsanalyse unter Variation dieser Modellparameter wird in den Abbildungen 10.5(a) und 10.5(b) dargestellt. Im Folgenden werden die Ergebnisse erörtert.

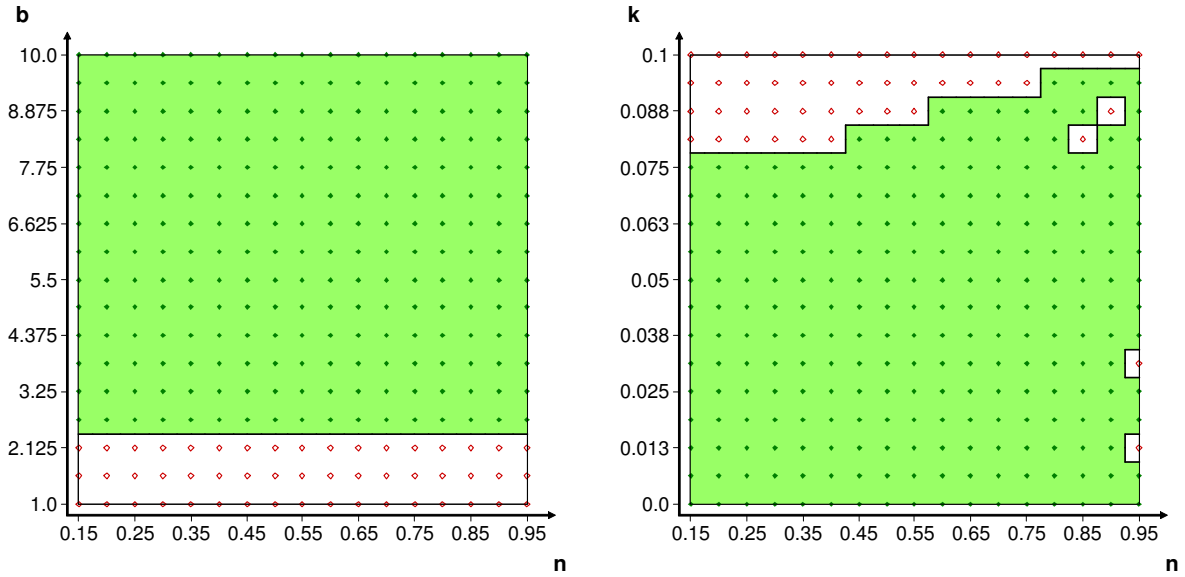

Abbildung 10.5: Sensibilitätsanalyse bezüglich des Anteils der normativen Einheiten n und **(a)** dem durchschnittlichen Nutzen-/Kostenverhältnis von Aktionen b und **(b)** dem durchschnittlichen Kostenverhältnis zwischen Nachrichten und Aktionen k

Die Variation von b zeigt, dass es eine untere Schranke b_{min} für das Verhältnis zwischen Nutzen und Kosten einer Aktionsausführung gibt. Wenn diese unterschritten wird, erreichen die Einheiten durch die Teilnahme am Informationssystem keinen positiven Individualnutzen und treten daher aus dem System aus. Eine genaue Analyse der Messergebnisse legt nahe, dass b_{min} im Bereich von $2,3$ liegt. Ein weiteres Ergebnis der Sensibilitätsanalyse ist, dass der Anteil normativer Einheiten keinen erkennbaren Einfluss auf diese Aussage besitzt.

Bei der Variation von k ergibt sich ein komplizierteres Bild. Es sind zwei Bereiche zu unterscheiden, in denen es zur Verletzung des Kriteriums der Existenzfähigkeit kommt. Zum einen ist dies vereinzelt bei einem hohen Anteil normativer Einheiten der Fall. Dies entspricht dem Ergebnis aus der Variation der Zahl der Transaktionsgelegenheiten. Zum anderen gibt es eine obere Schranke k_{max} für die maximal vertretbaren Kosten der Nachrichtenübermittlung. Wird sie überschritten, so kommt es auch hier zum Austritt von Benutzern aufgrund ihres negativen Individualnutzens. Die Höhe dieser oberen Schranke hängt vom Anteil normativer Einheiten ab. Je mehr Einheiten normativ sind, desto kooperativer ist das im Informationssystem vorherrschende Verhalten. Somit können bei einem hohen Anteil normativer Einheiten höhere Kosten verkraftet werden als bei einem niedrigen Anteil. Die Auswirkung dieses Effekts ist allerdings eher klein. Insgesamt lässt sich für die obere Schranke $k_{max} \approx 0,09$ festhalten.

Wie sind diese Ergebnisse für b und k zu interpretieren? Der Schlüssel zur Interpretation liegt im Verständnis davon, wie die Werte für die Schranken b_{min} und k_{max} zustande kommen. Hierzu müssen wir uns vor Augen halten, dass es nur dann zu Kooperation kommt, wenn sich die Einheiten zu entsprechenden Vertrauensentscheidungen durchringen können. Bei einem anfänglichen Typglauben von 50% muss gemäß Gleichung 6.12 für eine Einheit in der Risiko-Position somit folgende Bedingung erfüllt sein:

$$50\% \cdot \overline{p}_u(\gamma)^2 \cdot u(a_p) - c(a_o) - c_T > 0 \tag{10.5}$$

In diese Bedingung sind die Modellparameter b und k einzusetzen. Dies lässt sich bewerkstel-

ligen, indem wir $c(a_o)$ auf eins normieren. Als Folge davon ist $u(a_p) = b$. Außerdem erhalten wir $c_T = 3 \cdot k$, da eine Einheit für jeden ihrer drei Schritte in der Transaktion gemäß Abschnitt 7.2.2 eine Nachricht übermitteln muss. Für p_u setzen wir gemäß dem Entwurfspunkt aus Abschnitt A.1.1 den Wert $7,8\%$ ein. Somit erhalten wir die folgenden Bedingungen aufgelöst nach b und k:

$$\begin{aligned} b &> 2,35 \cdot (1 + 3 \cdot k) \\ k &< 0,142 \cdot b - 0,333 \end{aligned} \tag{10.6}$$

Unter Einsetzung der jeweiligen Standardbelegung von b und k erhalten wir aus diesen Bedingungen $b > 2,49$ und $k < 0,164$. Diese analytischen Abschätzung der Schranken b_{min} und k_{max} entspricht den gemessenen Werten ziemlich genau. Lediglich k_{max} scheint etwas überschätzt worden zu sein. Dies liegt daran, dass in dieser Herleitung das Nachrichtenaufkommen des Empfehlungssystems keine Berücksichtigung findet. Die dabei verursachten Kosten drücken auf den Individualnutzen der Einheiten und bewirken, dass es schon bei einem geringeren $k_{max} \approx 0,09$ zu Austritten kommt.

Welche Schlüsse lassen sich aus diesen Überlegungen ziehen? Damit es initial zur Bildung von Vertrauen zwischen den Einheiten kommen kann, bedarf es eines nicht zu geringen Verhältnisses zwischen dem Nutzen und den Kosten der Kooperation. Dieses ist umgekehrt proportional zum Systemglauben, den eine Einheit initial besitzt: Geht eine Einheit anfangs davon aus, dass der Anteil normativer Einheit 50% ist, so wird ein Nutzen-/Kostenverhältnis der Kooperation von ungefähr 2:1 benötigt, um die Einheit zum Aussprechen von Vertrauen und dem Eingehen in Transaktionen zu bewegen. Ist die anfängliche Einschätzung des Anteils normativer Einheiten höher (zum Beispiel 75%), so ist ein geringes Verhältnis ausreichend (im Beispiel $1,3$). Umgekehrt bedarf es eines umso höheren Verhältnisses, je misstrauischer die Benutzer dem Informationssystem entgegenstehen.

Anteil potentieller Transaktionspartner (p). Der Modellparameter p gibt an, wie hoch der Anteil der Einheiten, die bei einer Transaktionsgelegenheit als potentieller Partner in Frage kommen, im Durchschnitt ist. Abbildung 10.6(a) zeigt das Ergebnis der Sensibilitätsanalyse bezüglich dieses Modellparameters. Aus der Abbildung geht hervor, dass das Informationssystem unabhängig vom Anteil potentieller Transaktionspartner existenzfähig ist.

Um zu untersuchen, welchen Einfluss die Höhe des Anteils auf das Gesamtsystem ist, werden in Abbildung 10.6(b) ausgewählte Teile der Sensibilitätsanalyse dargestellt. Es handelt sich bei dieser Illustration um eine Gegenüberstellung zwischen dem Anteil p und den gemessenen Normativitätskosten. Die Messwerte ergeben sich bei einem Anteil normativer Einheiten von $n = 70\%$. Sie sind repräsentativ für andere Belegungen von n.

Die Abbildung zeigt, dass mit einem fallenden Anteil potentieller Transaktionspartner die Normativitätskosten steigen. Wodurch ergibt sich dieser Zusammenhang? Wenn bei einer Transaktionsgelegenheit nur eine Einheit als potentieller Transaktionspartner in Frage kommt, so reicht es zum Eingehen in die Transaktion aus, dass der erwartete Nutzen aus der Transaktion mit ihr positiv ist. Damit eine strategische Einheit als Transaktionspartner gewählt wird, muss sie somit lediglich als hinreichend normativ erscheinen. Die Situation ist aber eine andere, wenn mehrere Einheiten als potentielle Transaktionspartner zur Verfügung stehen. In diesem Fall wird diejenige Einheit gewählt, die am normativsten erscheint. Dadurch geraten Einheiten, die bereits an Konflikten beteiligt waren, in die Lage, selten oder gar nicht als Transaktionspartner gewählt zu werden. In Umgebungen mit einem hohen Anteil potentieller Transaktionspartner sind somit die Folgekosten von Betrugskosten besonders hoch. Ebendies schlägt sich im Verlauf der Normati-

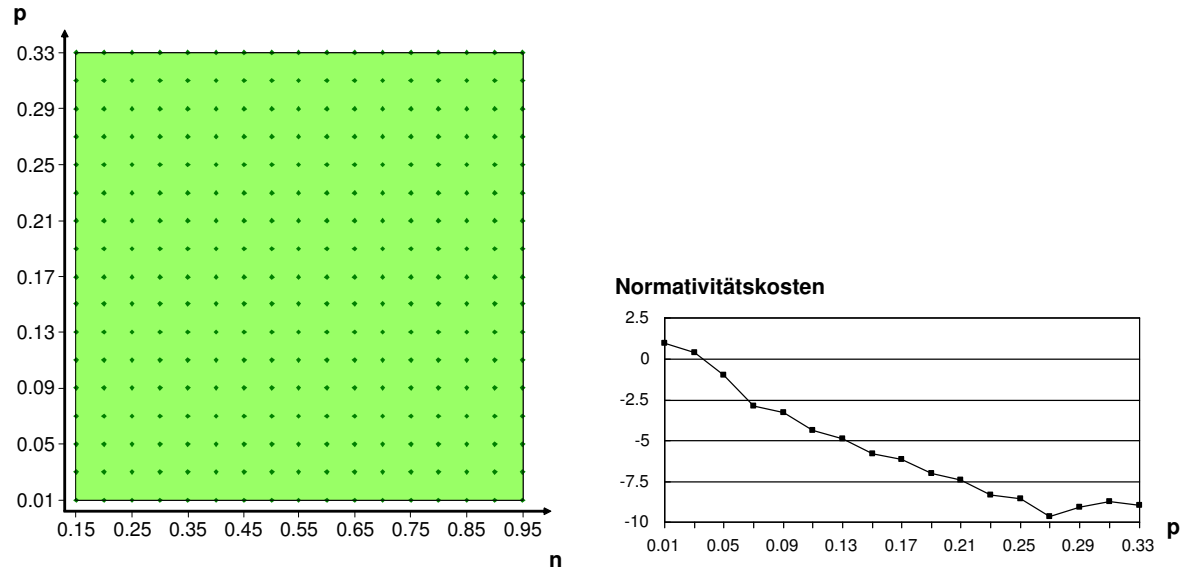

Abbildung 10.6: **(a)** Sensibilitätsanalyse bezüglich des Anteils der normativen Einheiten n und dem Anteil potentieller Transaktionspartner p; **(b)** Abhängigkeit der Normativitätskosten von p

vitätskosten nieder: Je höher dieser Anteil ist, desto schlechter schneiden strategische Einheiten im Vergleich zu normativen Einheiten ab.

Zahl der Einheiten im System (s). Abbildung 10.7(a) zeigt das Ergebnis der Sensibilitätsanalyse, in der die Zahl der Einheiten s variiert wird. Es ergibt sich ein ähnliches Bild, wie es sich zuvor bereits unter Variation der Zahl der Transaktionsgelegenheiten gezeichnet hat: Die Normativitätskosten sind unabhängig von der Systemgröße so gering, dass nur wenige Benutzer (bis zu 15%) zur Manipulation bereit sind.

Auch beim Modellparameter s stellen wir die Messergebnisse in Abbildung 10.7(b) detailliert dar, um seinen Einfluss auf das Gesamtsystem zu untersuchen. Wie zuvor sind dazu die Normativitätskosten bei einem Anteil normativer Einheiten von $n = 70\%$ aufgezeichnet. Der Verlauf ist repräsentativ für andere Belegungen von n.

Aus der Abbildung geht hervor, dass mit einer wachsenden Zahl von Einheiten die Normativitätskosten ansteigen. Dieser Zusammenhang erklärt aus der Betrachtung der Betrugskosten: Betrügt eine strategische Einheit in einem kleinen System, so fällt ihr Betrugsverhalten schneller auf als in einem großen System. Als Folge davon ist in einem großen System die Wahrscheinlichkeit höher, dass eine strategische Einheit trotz früheren Betrugsverhaltens von anderen Einheiten als Transaktionspartner angenommen wird. Insofern sinken die Betrugskosten mit der Zahl der Einheiten s. Allerdings müssen wir berücksichtigen, dass bei der Abbildung die anderen Modellparameter außer s unverändert belassen sind. Das würde unter anderem bedeuten, dass die Zahl der zu erwartenden Transaktionsgelegenheiten nicht von der Systemgröße abhängt. In Abschnitt 10.2.3 werden wir darlegen, warum diese Annahme unrealistisch ist. Eine abschließende Beurteilung der Skalierbarkeit des Informationssystems kann somit nur mit Hilfe einer weiteren Versuchsreihe getroffen werden. Dies wird in eben jenem Abschnitt getan.

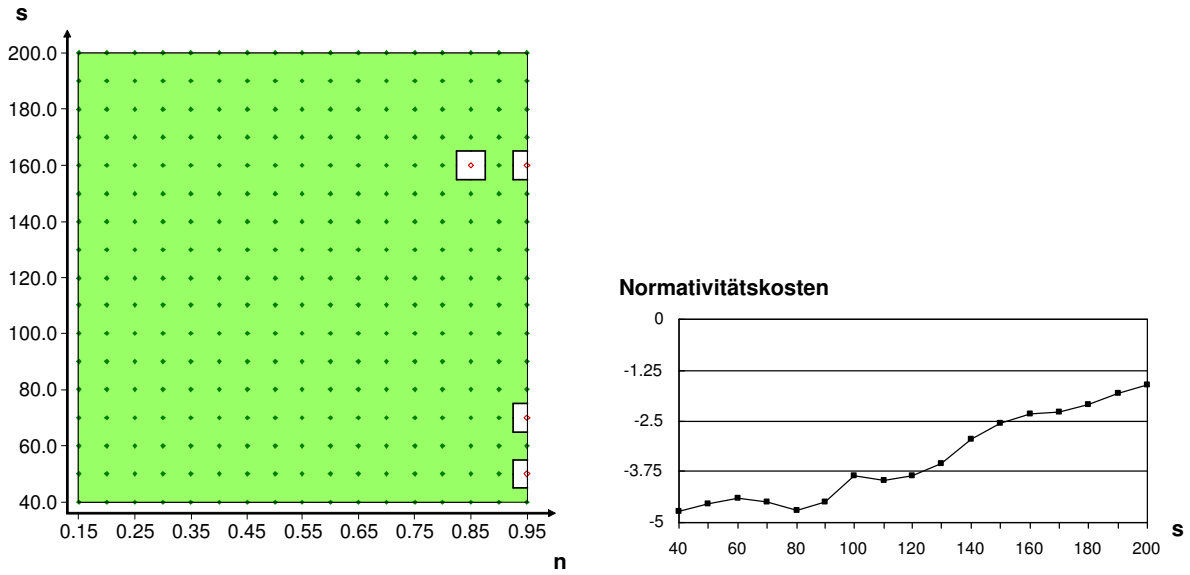

Abbildung 10.7: **(a)** Sensibilitätsanalyse bezüglich des Anteils der normativen Einheiten n und der Zahl der Einheiten s; **(b)** Abhängigkeit der Normativitätskosten von s

Wahrscheinlichkeit unbeabsichtigter Fehler bei der Aktionsausführung (u). Die Sensibilitätsanalyse aus Abbildung 10.8(a) zeigt, wie die Existenzfähigkeit des Informationssystems von der Wahrscheinlichkeit u abhängt, mit der während der Aktionsausführung sich ein unbeabsichtigter Fehler ereignet. Es lassen sich zwei Bereiche unterscheiden, in denen das Kriterium der Existenzfähigkeit nicht erfüllt wird. Zum einen ist dies vereinzelt bei einem hohen Anteil normativer Einheiten ($n \geq 90\%$) der Fall. Hierbei lässt sich kein eindeutiger Zusammenhang mit der Höhe von u feststellen. Zum anderen gibt es einen Bereich ab $u \geq 17,5\%$, in dem aufgrund einer zu hohen Fehlerrate der Individualnutzen der Benutzer zu gering wird und es in der Folge zu Austritten kommt. Dabei ist auffällig, dass ab einem mittleren Anteil normativer Einheiten eine höhere Fehlerrate (über 20%) verkraftet werden kann. Einen ähnlichen Zusammenhang haben wir bereits bei den Kosten der Nachrichtenübermittlung gesehen: Bei einem hohen Anteil normativer Einheiten herrscht kooperatives Verhalten im Gesamtsystem vor. Folglich sind höhere Fehlerraten eher tragbar.

Wie ist das Ergebnis der Sensibilitätsanalyse zu beurteilen? Zunächst ist zu berücksichtigen, dass in einer Transaktion beide Transaktionspartner je eine Aktion ausführen. Bei einer Fehlerrate von 20% pro Aktionsausführung werden somit maximal nur $(1 - 20\%)^2 = 64\%$ der Transaktionen, in denen beide Transaktionspartner kooperatives Verhalten beabsichtigen, erfolgreich zu Ende geführt. Wird diese Wahrscheinlichkeit unterschritten, so zeigen die Simulationsergebnisse, dass das Informationssystem nicht existenzfähig ist[5]. Hierbei ergibt sich ein enger Zusammenhang mit zu dem minimal erforderlichen Nutzen-/Kostenverhältnis der Kooperation: Das Nutzen-/Kostenverhältnis einer erfolgreichen Transaktion ist durchschnittlich $b/(1 + 3 \cdot k) \approx 3,3$. Allerdings sind nur 64% der Transaktionen, in denen Kooperation beidseitig beabsichtigt wird,

[5]Diese Aussage relativiert sich, wenn wir berücksichtigen, dass transiente Kommunikationsstörungen vom verwendeten Transportsystem abgefangen werden. Die Wahrscheinlichkeit u bezieht sich also darauf, dass der Kommunikationskanal unerwartet und dauerhaft während der Ausführung einer Aktion zusammenbricht. Die Wahrscheinlichkeit hierfür ist aber gemäß dem Systemmodell äußerst gering.

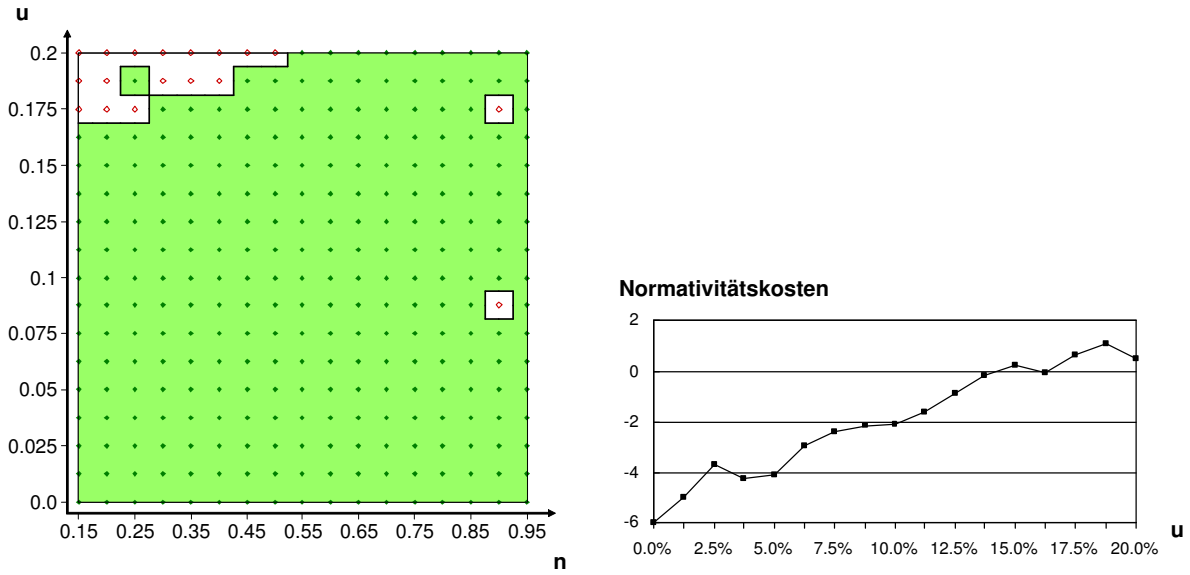

Abbildung 10.8: **(a)** Sensibilitätsanalyse bezüglich des Anteils der normativen Einheiten n und der Wahrscheinlichkeit eines unbeabsichtigten Fehlers bei der Aktionsausführung u; **(b)** Abhängigkeit der Normativitätskosten von u

erfolgreich. Auch in den übrigen Transaktionen müssen die Kosten der Kooperation getragen werden. Damit erhalten wir ein erwartetes Nutzen-/Kostenverhältnis von $64\% \cdot 3,3 + 36\% \cdot 0 = 2,1$. Dies entspricht dem Verhältnis von zwei-zu-eins, von dem wir bereits gesehen haben, dass es Voraussetzung für die Existenzfähigkeit des Informationssystems ist.

Abschließend gehen wir auf die Abhängigkeit der Normativitätskosten von der Fehlerrate u ein. Abbildung 10.8(b) zeigt ihren Verlauf. Wie in den vorangehenden Schaubildern sind die Messwerte für $n = 70\%$ abgebildet und für andere Belegungen von n repräsentativ. Aus der Abbildung geht hervor, dass mit wachsender Fehlerrate die Normativitätskosten steigen. Dieses Ergebnis ist nicht verwunderlich: Das Verhalten, das von normativen und strategischen Einheiten gezeigt wird, unterscheidet sich umso weniger, je wahrscheinlicher es zu unbeabsichtigtem Fehlverhalten kommt. Bei einer hohen Fehlerrate hat eine strategische Einheit somit höhere Chancen, dass ihr beabsichtigtes Betrugsverhalten als unbeabsichtigt eingestuft wird. Diese Einsicht lässt sich abstrakt auch so formulieren: Je höher die Fehlerrate desto geringer ist der Informationsgehalt von Transaktionserfahrungen für die Glaubensbildung.

Trotz dieser Einschränkungen zeigt der Verlauf der Normativitätskosten in der Abbildung, dass selbst bei hohen Fehlerraten strategische Einheiten kaum einen Vorteil gegenüber normativen Einheiten besitzen. Dies erklärt das Ergebnis der Sensibilitätsanalyse, dass auch bei hohen Fehlerraten nur ein kleiner Teil ($\leq 10\%$) der Benutzer (wenn überhaupt) zur Manipulation bereit ist.

10.2.3 Innere Abhängigkeit der Rahmenbedingungen

In den Versuchsreihen des vorigen Abschnitts wurde außer dem Anteil normativer Einheiten jeweils ein Modellparameter variiert. Damit konnten wir den Einfluss einzelner Modellparameter auf die Existenzfähigkeit des Informationssystems erfassen. In diesem Abschnitt ergänzen wir diese Betrachtungen, indem wir die Abhängigkeit der Modellparameter untereinander untersuchen.

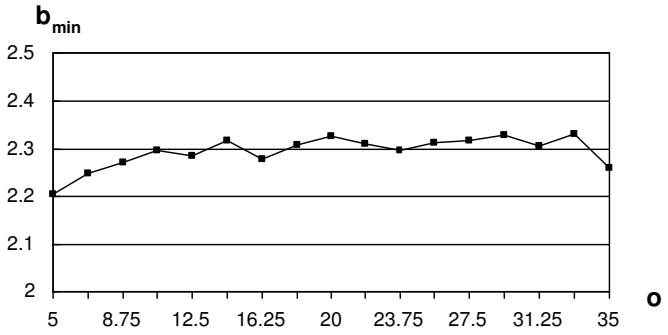

Abbildung 10.9: Minimal erforderliches Nutzen-/Kostenverhältnis der Aktionsausführung b_{min} in Abhängigkeit der durchschnittlichen Zahl der Transaktionsgelegenheiten pro Einheit o

Dazu werden Versuchsreihen vorgestellt, die gezielt jeweils ein Paar von Modellparametern variieren. Bei der Wahl der Paare werden wir davon geleitet, welche Abhängigkeiten eine gesonderte Betrachtung erfordern. Im Folgenden werden drei solcher Abhängigkeiten untersucht. Gemäß der Standardeinstellung gehen die jeweiligen Versuchsreihen von einem Anteil normativer Einheiten $n = 70\%$ aus, da die jeweiligen Sensibilitätsanalysen auf die Variation zweier Modellparameter beschränkt ist.

Erforderliches Nutzen-/Kostenverhältnis (b,o). Zunächst erweitern wir die Betrachtung zum minimal erforderlichen Nutzen-/Kostenverhältnis b_{min} der Aktionsausführung. Die Sensibilitätsanalyse aus dem vorigen Abschnitt zeigt, dass dieses b_{min} nicht vom Anteil normativer Einheiten abhängt. Im Folgenden untersuchen wir die Frage, ob die durchschnittliche Zahl der Transaktionsgelegenheiten pro Einheit o einen Einfluss auf dieses minimal erforderliche Verhältnis besitzt.

Die Sensibilitätsanalyse zur Variation von b und o gibt eine negative Antwort auf diese Frage. Wie in Abbildung 10.5(a) zeigt sich das Bild, dass die untere Schranke bei b_{min} nicht von o abhängt. Abbildung 10.9 gibt genauen Aufschluss darüber, wie hoch b_{min} in Abhängigkeit von o ist. Wir erhalten durchweg $b_{min} \approx 2,3$ mit nur sehr geringen Abweichungen. Damit ist b_{min} eine Größe, die weder vom Anteil normativer Einheiten noch von der Zahl der Transaktionsgelegenheiten abhängt.

Welche Aussage können wir daraus ableiten? Die analytische Abschätzung von b_{min} im vorangegangenen Abschnitt benötigt keine Information über die Größe von n oder o, um zu einer guten Annährung zum gemessenen Wert. Die Sensibilitätsanalysen bestätigen somit die Präambel der Abschätzung, dass sich der Wert von b_{min} alleine aus der Notwendigkeit zur initialen Bildung von Vertrauen ergibt. Die Versuchsreihe untermauert also die Aussagen, die wir im vorangegangenen Abschnitt zum Verhältnis zwischen b_{min} und dem initialen Systemglauben gemacht haben.

Realitätsnahe Variation der Zahl der Einheiten (s,o). In Abschnitt 10.2.2 wurde die Zahl der Einheiten s nur in Abhängigkeit des Anteils normativer Einheiten variiert. Insbesondere wurde nicht der Zusammenhang mit der durchschnittlichen Zahl der Transaktionsgelegenheiten pro Einheit o untersucht. Dies ist insofern problematisch, als die Systemgröße unter realistischen Bedingungen sehr wohl einen Einfluss auf die Zahl der Transaktionsgelegenheiten besitzt. Dies wird aus der Szenariobeschreibung deutlich: Wenn das Informationssystem nur aus den Geräten

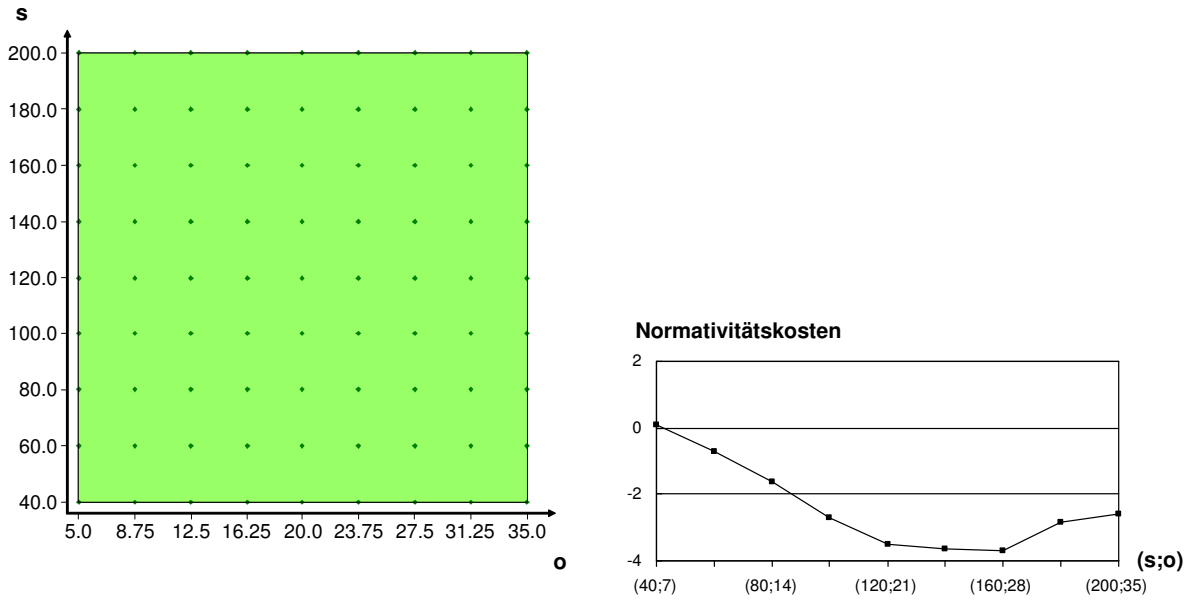

Abbildung 10.10: **(a)** Sensibilitätsanalyse bezüglich der durchschnittlichen Zahl der Transaktionsgelegenheiten pro Einheit o und der Zahl der Einheiten s; **(b)** Abhängigkeit der Normativitätskosten von den als proportional angenommenen Größen s und o

der sechs beschriebenen Benutzer besteht, so gibt es für Annas Vorlesungsmitschrieb nur zwei Interessenten. Auf der anderen Seite sind für Anna nur zwei der angebotenen Informationen oder Informationsdienste von Interesse. Die Situation ändert sich, wenn sehr viel mehr Benutzer am Informationssystem teilnehmen. In diesem Fall stünden Anna auch die Vorlesungsmitschriebe ihrer Kommilitonen und andere Unterlagen und Informationen, die für sie von Interesse sind, zur Verfügung. Auf der anderen Seite gäbe es auch mehr Interessenten am Vorlesungsmitschrieb, den sie erstellt. Wir erhalten somit, dass ein direkter Zusammenhang zwischen der Systemgröße s und der Zahl der Transaktionsgelegenheiten pro Einheit o besteht. Für Betrachtungen der Skalierbarkeit ist dieser Zusammenhang zu berücksichtigen.

Basierend auf dieser Überlegung wird eine Sensibilitätsanalyse unter Variation von s und o durchgeführt. Das Ergebnis zeigt Abbildung 10.10(a). Unabhängig von s und o ist das Informationssystem existenzfähig. Dieses Resultat ist nicht verwunderlich, da wir es für $n = 70\%$ im vorangegangenen Abschnitt bereits zum Teil erhalten haben. Darüber hinaus untersuchen wir den Verlauf der Normativitätskosten. Er ist in Abbildung 10.10(b) dargestellt. Dabei wird nicht nur wie in Abbildung 10.7(b) die Zahl der Einheiten s variiert. Zudem wird auch die Zahl der Transaktionsgelegenheiten pro Einheit o in einem proportionalen Zusammenhang zu s variiert. Dies ist erforderlich, um den zuvor gezeigten Einfluss von s auf o berücksichtigen zu können. Der Verlauf zeigt, dass mit einer wachsenden Zahl von Einheiten die Normativitätskosten fallen. Bei Werten von $s \approx 180$ steigen die Normativitätskosten wieder ein wenig. Sie bleiben aber weit unter denen für geringe s. So schneiden zum Beispiel strategische Einheiten in einem System der Größe 40 weitaus besser ab als in einem System der Größe 200.

Welchen Schluss können wir aus diesem Verlauf ziehen? In der Evaluation des Gesamtsystems werden vorwiegend Informationssysteme simuliert, die aus 100 Einheiten bestehen. Die Variation dieser Zahl hat gezeigt, dass auch von größeren Informationssystemen ($s \approx 200$) die Existenzfähig-

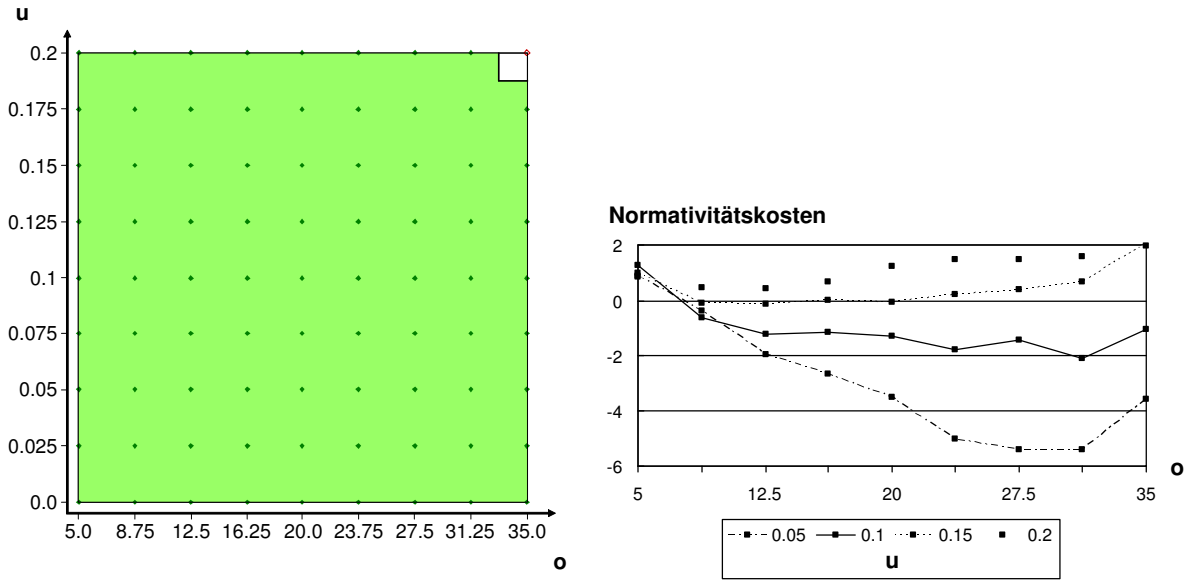

Abbildung 10.11: **(a)** Sensibilitätsanalyse bezüglich der durchschnittlichen Zahl der Transaktionsgelegenheiten pro Einheit o und der Wahrscheinlichkeit für unbeabsichtigte Fehler der Aktionsausführung u; **(b)** Abhängigkeit der Normativitätskosten von o und u

keit gesichert ist. Zudem legen die Messwerte die Vermutung nahe, dass auch deutlich größere Informationssysteme existenzfähig sind. Insofern wird mit dieser Versuchsreihe gezeigt, dass der eigene Ansatz der verteilten Vertrauensbildung für unterschiedliche Systemgrößen erfolgreich einsetzbar ist und somit skaliert.

Vertretbare Fehlerraten der Aktionsausführung (u,o). In Abschnitt 10.2.2 haben wir gesehen, dass allzu hohe Wahrscheinlichkeiten u für unbeabsichtigte Fehler der Aktionsausführung die Existenzfähigkeit des Informationssystems gefährden. Im Folgenden untersuchen wir, wie diese Aussage von der durchschnittlichen Zahl der Transaktionsgelegenheiten pro Einheit o abhängt. Insbesondere ist herauszufinden, ob hohe Fehlerraten durch eine Vielzahl von Transaktionsgelegenheiten eher verkraftet werden kann.

Abbildung 10.11(a) zeigt das Ergebnis der Sensibilitätsanalyse, in der u und o variiert werden. Unabhängig von der Wahl der beiden Modellparameter ist das Informationssystem existenzfähig (einzige Ausnahme $u = 20\%$ und $o = 35$). Dies ergibt sich unmittelbar aus dem Resultat der Abbildung 10.8(a), dass bei einem Anteil normativer Einheiten $n = 70\%$ eine Fehlerrate von $u = 20\%$ vertretbar ist. Im Rahmen dieser Versuchsreihe ist es daher aufschlussreicher, die Normativitätskosten in Abhängigkeit von u und o zu untersuchen. Abbildung 10.11(b) zeigt den Verlauf abhängig von o, der sich für unterschiedliche Belegungen von u ergibt. Wir erhalten den folgenden Zusammenhang: Für die Standardbelegung $u = 5\%$ fallen die Normativitätskosten mit steigendem o. Ausnahme ist lediglich der Bereich $o \approx 30$, in dem die strategischen Einheiten gemäß der Gegenstrategie DISTRUSTDISTRIBUTOR in die Betrugsphase wechseln. Einen ähnlichen Verlauf erhalten wir auch für $u = 10\%$, wenngleich die Normativitätskosten nur wenig mit steigendem o fallen. Bei höheren Fehlerraten $u \geq 15\%$ lässt sich jedoch kein solcher Zusammenhang zwischen o und den Normativitätskosten erkennen. Bei Fehlerraten um $u \approx 20\%$ steigen die Normativitätskosten sogar leicht mit wachsendem o.

Wie sind diese Verläufe zu interpretieren? Unter realistischen Annahmen der Fehlerrate ($u \leq 5\%$) schneiden normative Einheiten gegenüber strategischen Einheiten umso besser ab, je mehr Transaktionsgelegenheiten sie erhalten. Dies liegt daran, dass die strategischen Einheiten als Folge ihres Betrugsverhaltens immer weniger als Transaktionspartner angenommen werden. Wenn die Fehlerrate jedoch sehr hoch wird ($u \geq 15\%$), ändert sich die Situation. Es ist zwar weiterhin so, dass strategische Einheiten durch ihr beabsichtigtes Betrugsverhalten seltener als normative Einheiten als Transaktionspartner gewählt werden. Jedoch sind die Auswirkungen dieses Effekts geringer. Dies liegt daran, dass die Wahrscheinlichkeit unbeabsichtigten Betrugsverhaltens sehr viel höher ist und somit betrügende Einheiten weniger abgewertet werden. In der Summe sind daher die Betrugskosten zu gering, als dass Betrugsverhalten nachteilig ist. Dies zeigt sich insbesondere im Verlauf für $u = 20\%$.

Wir fassen zusammen, dass sehr hohe Fehlerraten die Existenzfähigkeit des Informationssystems nicht nur im Hinblick auf mögliche Austritte der Benutzer gefährden. Zudem kommt es zu einer Verringerung der Betrugskosten, die die Manipulation vorteilhafter als zuvor erscheinen lässt. Allerdings tritt diese Erscheinung erst bei unrealistisch hohen Fehlerraten ($u \geq 20\%$) zu Tage. Außerdem bleibt ihre Wirkung eher klein, da selbst bei $u = 20\%$ die Normativitätskosten für alle Variationen von o sehr gering sind.

10.2.4 Einfluss menschlicher Eigenschaften und Präferenzen

Die menschlichen Eigenschaften und Präferenzen werden durch zwei Modellparameter festgelegt: Die Wahrnehmung der Rahmenbedingungen durch manipulationswillige Benutzer und die Verteilung der Manipulationskosten. Im Folgenden ergänzen wir die Ergebnisse der vorangegangenen Abschnitte um Versuchsreihen, in denen diese beiden Modellparameter variiert werden.

Wahrnehmung der Rahmenbedingungen. Manipulationswillige Benutzer wählen diejenige manipulierte Version, die ihnen bei Betrachtung der Rahmenbedingungen am geeignetsten erscheint. Gemäß Abschnitt 10.1.2 gehen wir realistischerweise davon aus, dass die Benutzer die Rahmenbedingungen nicht genau kennen. Genauer gesagt kommt es zur Unter- beziehungsweise Überschätzung einzelner Aspekte der Rahmenbedingungen. Dies wird durch die Faktoren l und r festgehalten. Bisher sind wir davon ausgegangen, dass die Benutzer den Anteil normativer Einheiten fast genau einschätzen können ($l_n = 0,8$ und $r_n = 1,2$) und eine ungefähre Vorstellung von der Zahl der Transaktionsgelegenheiten haben ($l_o = 0,5$ und $l_o = 2,0$). Es stellt sich somit die Frage, ob das Informationssystem existenzfähig bleibt, wenn die Benutzer die Rahmenbedingungen präziser wahrnehmen können.

Diese Frage untersuchen wir, indem wir den Extremfall annehmen, dass die Rahmenbedingungen von den Benutzern genau gekannt werden ($l_n = 1 = r_n$ und $l_o = 1 = r_o$). Wie in Abschnitt 10.1.2 gezeigt ist dieser Fall zwar sehr unrealistisch. Wir erhalten dadurch jedoch Aufschluss darüber, wie sehr die Existenzfähigkeit des Informationssystems durch allzu gute Kenntnis der Rahmenbedingungen seitens der Benutzer gefährdet werden kann.

Abbildung 10.12(a) zeigt die Ergebnisse einer Sensibilitätsanalyse, bei der n und o variiert werden. Diese Versuchsreihe geht davon aus, dass die Rahmenbedingungen präzise wahrgenommen werden. Aus der Abbildung geht dasselbe Ergebnis hervor, das wir bereits für die unpräzise Wahrnehmung in Abbildung 10.4(a) erhalten haben: Das Informationssystem ist unter allen Rahmenbedingungen existenzfähig. Lediglich ein kleiner Teil ($< 10\%$) inhärent manipulationsfreudiger Benutzer entscheidet sich unter Umständen zur Manipulation.

Um den Unterschied zwischen präziser und unpräziser Wahrnehmung erfassen zu können,

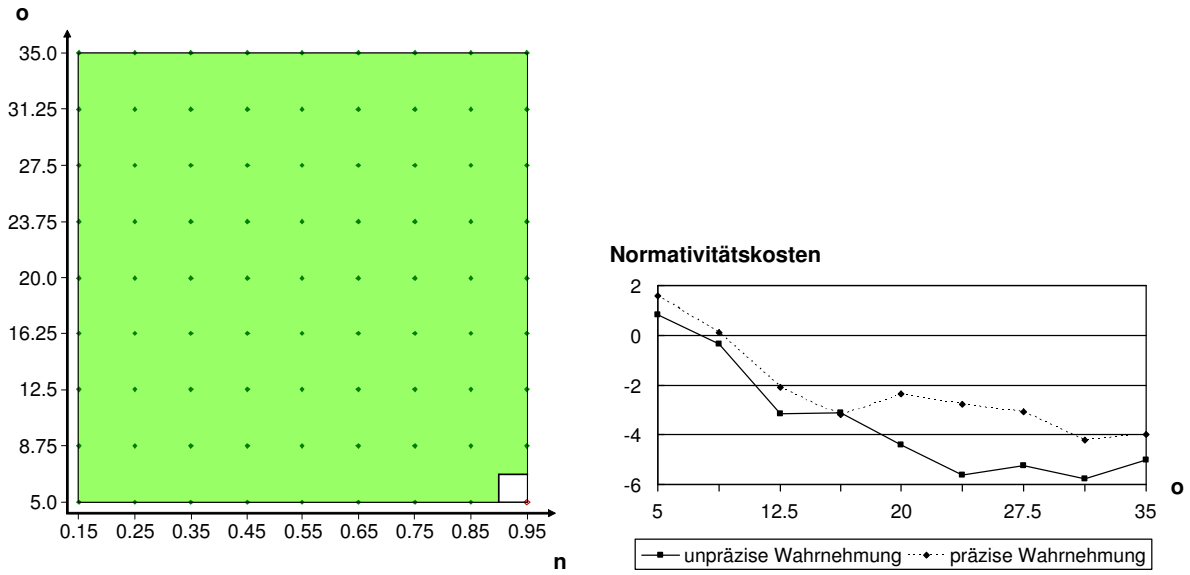

Abbildung 10.12: **(a)** Sensibilitätsanalyse bezüglich des Anteils der normativen Einheiten n und der durchschnittlichen Zahl der Transaktionsgelegenheiten pro Einheit o bei präziser Wahrnehmung der Rahmenbedingungen; **(b)** Vergleich der Normativitätskosten unter präziser und unpräziser Wahrnehmung

bedarf es also einer detaillierteren Darstellungsweise. Dies wird in Abbildung 10.12(b) umgesetzt, indem die Normativitätskosten unter präziser und unpräziser Wahrnehmung verglichen werden. Dabei gehen wir von einem Anteil normativer Einheiten $n = 75\%$ aus. Die erhaltenen Messergebnisse sind repräsentativ für andere Belegungen von n. Es zeigt sich, dass die Präzision der Wahrnehmung sehr wohl einen Einfluss auf das Abschneiden der strategischen Einheiten hat. Wenn manipulationswillige Benutzer die Rahmenbedingungen präzise wahrnehmen können, sind sie in der Lage, die jeweils geeignetste Gegenstrategie zu wählen. Dies hat zur Folge, dass die strategischen Einheiten besser abschneiden, als wenn die Rahmenbedingungen nur unpräzise wahrgenommen werden können. Allerdings ist der sich ergebende Unterschied sehr gering.

Welche Aussage können wir aufgrund dieser Versuchsreihe treffen? Es wirkt sich zwar auf die Ergebnisse aus, wie präzise die Benutzer die Rahmenbedingungen wahrnehmen. Der Einfluss davon ist jedoch so gering, dass sich für die Betrachtung der Existenzfähigkeit des Informationssystems keine Änderungen ergeben. Eben hierin liegt der Grund, warum das Ergebnis der Sensibilitätsanalyse unter präziser Wahrnehmung aus Abbildung 10.4(a) nicht von dem unter unpräziser Wahrnehmung abweicht. Folglich gelten die Ergebnisse aus den vorangegangenen Abschnitten auch für den unrealistischen Fall, dass die Benutzer die Rahmenbedingungen genau kennen.

Verteilung der Manipulationskosten. Die Verteilungsfunktion $F(m)$ hält fest, welcher Anteil der Benutzer Manipulationskosten von bis zu m besitzt. Bislang sind wir von einer Verteilung gemäß Abbildung 10.3 ausgegangen. Im Folgenden untersuchen wir, welche Auswirkungen sich ergeben, wenn die Manipulationskosten auf eine andere Weise verteilt sind.

Abbildung 10.13 zeigt zwei alternative Verteilungen der Manipulationskosten, die wie folgt

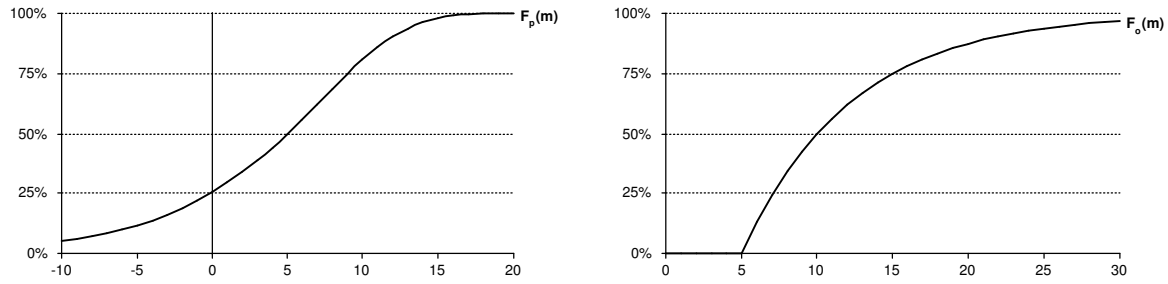

Abbildung 10.13: Alternative Verteilungen der Manipulationskosten

definiert sind:

$$\begin{aligned} F_p(m) &= 1 - (\frac{1}{2})^{2^{(m-5)/4}} \\ F_o(m) &= 1 - (\frac{1}{2})^{\frac{m-5}{5}} \end{aligned} \tag{10.7}$$

Der Verlauf der Verteilungsfunktionen ist folgendermaßen zu interpretieren:

- *Pessimistische Abschätzung:* Die Funktion $F_p(m)$ stellt eine äußerst pessimistische Abschätzung der Verteilung der Manipulationskosten dar. Sie geht nicht nur davon aus, dass für 25% der Benutzer das Überwinden der technischen und rechtlichen Hindernisse keinen Aufwand verursacht. Zudem wird für diese Benutzer angenommen, dass sie sich auch dann zur Manipulation entscheiden, wenn dies für sie Nachteile mit sich bringt. In der Funktion kommt dieser Sachverhalt durch $F_p(0) = 25\%$ zum Ausdruck. Es handelt sich bei diesen 25% der Benutzer um *bösartige* Studenten, die manipulierte Versionen der Systemsoftware benutzen wollen, auch wenn dies nachteilig ist. Ein weiterer Anteil von 25% der Studenten besitzt Manipulationskosten unter 5. Um die Ausführung von fünf Aktionen hinzunehmen, ziehen es diese Benutzer somit vor, eine manipulierte Version zu erstellen oder zu übernehmen und die dabei anfallenden substantiellen Hindernisse zu überwinden. Insgesamt zeigt diese Besprechung, dass es sich bei $F_p(m)$ um eine äußerst pessimistische Abschätzung der Manipulationskosten handelt.

- *Optimistische Abschätzung:* Die Verteilungsfunktion $F_o(m)$ ist eine optimistischere Abschätzung der Manipulationskosten. Sie geht realistischerweise davon aus, dass für jeden Student, auch wenn er noch so technisch versiert ist, das Überwinden der technischen Hindernisse für die Manipulation einen gewissen Aufwand verursacht. Dies kommt in der Funktion dadurch zum Ausdruck, dass jeder Benutzer mindestens die Manipulationskosten von 5 besitzt ($F_o(5) = 0$).

Wie wirkt es sich auf die Existenzfähigkeit des Informationssystems aus, wenn wir die pessimistische Abschätzung der Manipulationskosten annehmen? Abbildung 10.14(a) zeigt das Ergebnis der Sensibilitätsanalyse unter Variation von n und o und unter Annahme von $F_p(m)$. Aus der Abbildung geht hervor, dass nunmehr nur noch Informationssysteme mit einem Anteil normativer Einheiten unter 70% mit Sicherheit existenzfähig sind. Dieses Ergebnis ist nicht verwunderlich, da wir bei der Verteilung der Manipulationskosten $F_p(m)$ angenommen haben, dass 25% der Benutzer bösartig sind. Die Prinzipale der strategischen Einheiten rekrutieren sich somit zum

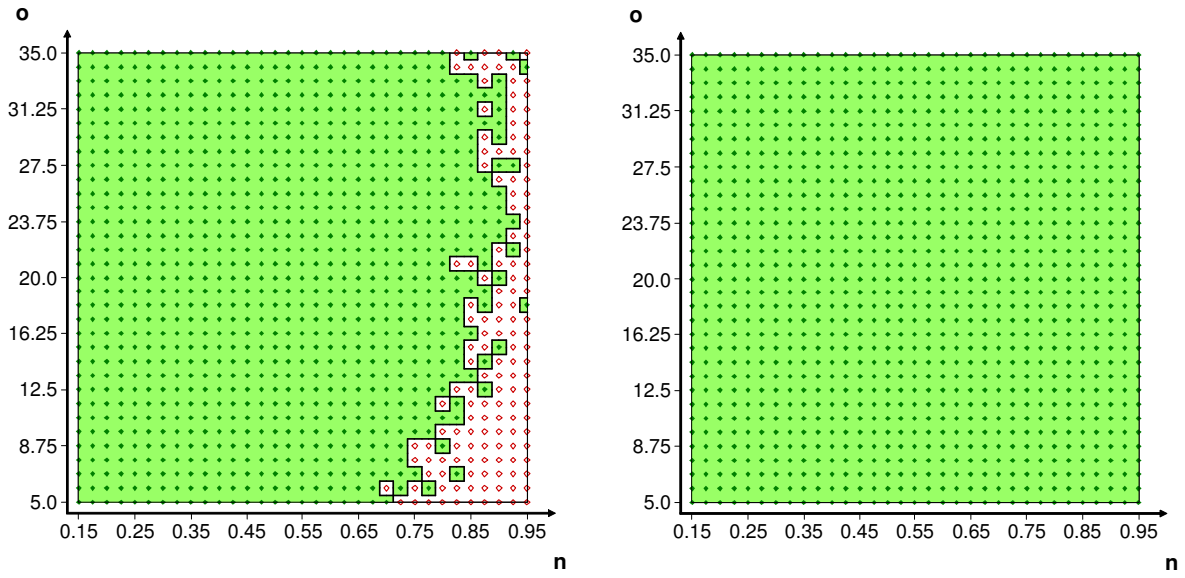

Abbildung 10.14: Sensibilitätsanalyse bezüglich des Anteils der normativen Einheiten n und der durchschnittlichen Zahl der Transaktionsgelegenheiten pro Einheit o, wobei für die Manipulationskosten gemäß **(a)** $F_p(m)$ und **(b)** $F_o(m)$ verteilt sind

überwiegenden Teil aus dieser Benutzermenge. Weiterhin zeigt die Abbildung, dass für $o \geq 12,5$ nur maximal 20% der Benutzer zur Manipulation bereit sind. Je nach Ausprägung von o liegt dieser Wert sogar bei bis zu 10%. Das bedeutet, dass die Vorteile für die Verwendung der originalen Systemsoftware derart groß sind, dass selbst einige der bösartigen Benutzer auf die Manipulation verzichten. Der Anteil der bösartigen Benutzer, die die Verwendung der originalen Systemsoftware vorzieht, beträgt bis zu 50%. Damit setzt sich der Systementwurf in besonders nachhaltiger Weise durch.

Abbildung 10.14(b) zeigt das Ergebnis der Sensibilitätsanalyse unter der Annahme, dass die optimistische Abschätzung der Manipulationskosten $F_o(m)$ zutrifft. Aus der Abbildung geht hervor, dass für alle betrachteten Rahmenbedingungen das Informationssystem existenzfähig ist. Der Anteil der Benutzer, die sich zur Manipulation entscheiden, liegt somit mit Sicherheit unter 5%. Dieses Ergebnis ist nicht überraschend, wenn wir uns vor Augen halten, dass die Normativitätskosten bei keiner Messung den Wert 5 überschreiten. Da gemäß $F_o(m)$ alle Benutzer mindestens die Manipulationskosten von 5 besitzen, ist das Kriterium der Existenzfähigkeit bei jeder Messung erfüllt. Dieses Ergebnis ist insofern von Wichtigkeit, als wir gesehen haben, dass die optimistische Abschätzung der Manipulationskosten eindeutig realitätsnäher als die Abschätzungen $F(m)$ und $F_p(m)$ ist.

Welche Schlussfolgerungen können wir aus diesen Versuchsreihen ziehen? Unterschiedliche Ausprägungen der Manipulationskosten haben zwar einen Einfluss darauf, welcher Anteil der Benutzer sich zur Manipulation entscheiden. An den grundsätzlichen Aussagen zur Existenzfähigkeit des Informationssystems ändert dies jedoch nichts. Entscheidend für den Anteil der manipulationswilligen Benutzer ist, welcher Anteil der Benutzer bösartig ist. Bei diesen handelt es sich um Benutzer, die sich zur Manipulation entscheiden, obwohl dies zu ihrem Nachteil ist. Wenn es keine solchen bösartigen Benutzer gibt und die Hindernisse zur Manipulation realistischerweise jedem Benutzer gewisse Kosten verursachen, erhalten wir sogar, dass keiner der Benutzer sich zur Ma-

nipulation entschließt. Damit ist der Systementwurf gemäß der Zielvorgabe aus Abschnitt 5.4.1 überaus erfolgreich.

10.2.5 Fazit

Das Grundproblem von Informationssystemen wie demjenigen des Campus-Szenarios ist, dass die Benutzer dem Systementwerfer gegenüber autonom sind und somit mit einer manipulierten Version der Systemsoftware am Informationssystem teilnehmen können. Durch das Betrugsverhalten der Einheiten dieser Benutzer kommt es zur Degeneration des Informationssystems und letztlich zum Aufhören seiner Existenz. Um dennoch die Existenzfähigkeit des Informationssystems gewährleisten zu können, ist ein Ansatz zur effektiven Betrugsvermeidung vonnöten. Die einzige Möglichkeit hierfür liegt in der sozialen Kontrolle der Einheiten untereinander, da das Informationssystem selbstorganisierend ist. Im Teil II dieser Arbeit wurde eine verteilte Vertrauensbildung entworfen, durch die diese soziale Kontrolle umgesetzt wird. Die Aufgabe der simulativen Evaluation liegt darin zu überprüfen, ob durch diesen Entwurf die These von der Existenzfähigkeit des Informationssystems validiert werden kann. Die Ergebnisse der Evaluation führen uns zu folgendem Schluss:

Validierung der These:

- Der eigene Entwurf der verteilten Vertrauensbildung sorgt für die Existenzfähigkeit von Informationssystemen, die wie im Campus-Szenario vollständig auf die Geräte der autonomen Teilnehmer verteilt sind.
- Unsere Vision von solchen selbstorganisierenden Informationssystemen lässt sich somit realisieren.
- Nur ein sehr geringer Anteil inhärent manipulationsfreudiger Benutzer entscheidet sich zur Verwendung manipulierter Versionen der Systemsoftware.
- Diese Benutzer können sich aber dadurch weder einen nennenswerten Vorteil verschaffen noch dem Gesamtsystem Schaden zufügen.

Dieses Ergebnis haben wir nicht nur für Rahmenbedingungen erhalten, die für das Informationssystem des Campus-Szenarios zu erwarten sind. Die Variation der Modellparameter hat gezeigt, dass für die allermeisten denkbaren Rahmenbedingungen die These von der Existenzfähigkeit des Informationssystems validiert werden kann.

Woran liegt dieses in seiner Klarheit doch überraschende Ergebnis? Durch den Einsatz der *verteilten Vertrauensbildung* wird für die soziale Kontrolle der Einheiten untereinander gesorgt. Diese bringt mit sich, dass strategische Einheiten sich kaum einen Vorteil gegenüber normativen Einheiten verschaffen können. Die Vorteile, die sich aus der Erstellung oder Übernahme manipulierter Versionen der Systemsoftware ergeben, sind somit minimal. Auf der anderen Seite sind für die Manipulation *technische und rechtliche Hindernisse* zu überwinden. Dies verursacht den Benutzern nicht vernachlässigbare Kosten, die den minimalen Manipulationsvorteil deutlich in den Schatten stellen. Aus diesem Grund entscheiden sich fast alle Benutzer zur Verwendung der originalen Systemsoftware. Dies führt zum Ergebnis, dass sich der Systementwurf durchsetzt und in der Folge das Informationssystem existenzfähig ist. Zusammenfassend erhalten wir dieses Resultat somit aufgrund des Einsatzes der verteilten Vertrauensbildung und der Existenz der Hindernisse zur Manipulation.

Die Evaluation des Gesamtsystems hat gezeigt, welche Voraussetzungen erfüllt sein müssen, damit die Existenzfähigkeit des Informationssystems zugesichert werden kann. Grundvoraussetzung ist ein hinreichend hohes Verhältnis zwischen dem Nutzen und den Kosten der Kooperation, damit es zwischen den Einheiten zur initialen Bildung von Vertrauen kommen kann. Wenn die Einheiten nicht allzu misstrauisch sind, ist ein Verhältnis von zwei-zu-eins hierfür ausreichend. Eine weitere Voraussetzung ist, dass die Kooperation nicht zu stark vom Umfeld des Informationssystems gestört wird. Entscheidend hierfür ist die Wahrscheinlichkeit, dass der Kommunikationskanal während einer Transaktion unvorhersehbar und dauerhaft zusammenbricht. Ist diese Wahrscheinlichkeit zu hoch, so werden die Einheiten zu oft um die Früchte der Kooperation gebracht und ihre Prinzipale wenden sich vom Informationssystem ab. Entscheidend ist hierbei wiederum, dass unter Berücksichtigung der Kosten, die durch diese Störungen verursacht werden, das erwartete Nutzen-/Kostenverhältnis der Kooperation weiterhin oberhalb von zwei-zu-eins liegt. Ein geringes Problem stellen hingegen die inhärent manipulationsfreudigen und bösartigen Benutzer dar, die sich als einzige Benutzer unter gewissen Bedingungen zur Manipulation entscheiden. Solange diese Benutzergruppe eine Minderheit darstellt, setzt sich der Systementwurf zu einem hohen Grade durch. Selbst wenn 25% der Benutzer bösartig sind, ist die Existenzfähigkeit des Informationssystems nicht gefährdet. Die Evaluation des Gesamtsystems hat zeigen können, dass sogar viele der bösartigen Benutzer auf die Manipulation verzichten und mit der originalen Systemsoftware am Informationssystem teilnehmen. Damit setzt sich der Systementwurf in besonders nachhaltiger Weise durch.

10.3 Zusammenfassung

In der simulativen Evaluation des eigenen Ansatzes ist herauszufinden, unter welchen Rahmenbedingungen die Existenzfähigkeit des Informationssystems gewährleistet ist. Zu diesem Zweck wurde in diesem Kapitel die zuvor entwickelte Evaluationsmethodik eingesetzt.

Im ersten Schritt des Evaluationsprozesses ging es darum zu antizipieren, wie manipulationswillige menschliche Benutzer auf den normativen Systementwurf reagieren. Zunächst wurden hierfür mit Hilfe des Interaktiven Kooperationsturniers viel versprechende Gegenstrategien *gefunden*, deren Einsatz von manipulationswilligen Benutzern zu erwarten ist. Die Gegenstrategien wurden in zwei Gruppen eingeteilt: Die einfachen Gegenstrategien zeichnen sich durch ihre kurzfristig ausgerichteten Betrugskriterien aus, während die komplexen Gegenstrategien einen großen Wert auf die initiale Bildung von Vertrauen legen und somit längerfristig angelegt sind. Die gefundenen Gegenstrategien wurden anschließend mit Hilfe des Simulativen Kooperationsturniers *bewertet*. Aus den Simulationsergebnissen ging hervor, unter welchen Rahmenbedingungen welche Gegenstrategie am erfolgreichsten ist. Das Resultat wurde für den Fall erweitert, dass die manipulationswilligen Benutzer die Rahmenbedingungen des Informationssystems nur unpräzise wahrnehmen können. Als Ergebnis erhielten wir eine Aussage darüber, in welchen Rahmenbedingungen welche Gegenstrategien realistischerweise zu erwarten sind.

Anschließend wurde die eigentliche *Evaluation des Gesamtsystems* mit Hilfe des Simulativen Kooperationsturniers durchgeführt. Dafür gingen wir zunächst auf die Standardeinstellung der Modellparameter ein, die den Ausgang für gezielte Variationen der Rahmenbedingungen bildet. Die Darstellung und Interpretation der Versuchsreihen wurde nach unterschiedlichen Gesichtspunkten gegliedert: Zunächst untersuchten wir durch die Variation des Anteils normativer Einheiten, welche Populationsstruktur in Abhängigkeit der Rahmenbedingungen in einem existenzfähigen Informationssystem zu erwarten ist. Anschließend wurden Versuchsreihen vorgestellt,

die gezielt auf die Abhängigkeiten zwischen einzelnen Aspekten der Rahmenbedingungen eingehen. Weiterhin wurde der Einfluss der menschlichen Eigenschaften und Präferenzen untersucht, indem unterschiedliche Modelle für die menschliche Wahrnehmung der Rahmenbedingungen und die Verteilung der Manipulationskosten angenommen wurden. Die Ergebnisse der Gesamtevaluation wurden in einem Fazit zusammengefasst. Darin wurde gezeigt, dass unsere These von der Existenzfähigkeit der Informationssysteme, die wie im Campus-Szenario vollständig auf die Geräte autonomer Teilnehmer verteilt sind, zutrifft.

Teil IV

Abschließende Betrachtungen

Kapitel 11

Zusammenfassung

In diesem Kapitel fassen wir die Vorgehensweise und Ergebnisse der ersten drei Teile dieser Arbeit zusammen. Die Ausgangspunkte (Teil I) werden in Abschnitt 11.1 besprochen. Anschließend wenden wir uns in Abschnitt 11.2 dem Entwurf (Teil II) zu. Mit dessen Evaluation (Teil III) der Arbeit befassen wir uns in Abschnitt 11.3. Den Abschluss dieses Kapitels bildet eine Zusammenstellung der wesentlichen Beiträge dieser Dissertation zum Stand der Forschung in Abschnitt 11.4.

11.1 Ausgangspunkt

Durch die Verbreitung von Informationsgeräten liegt es für menschliche Benutzer nahe, die Informationen, über die sie verfügen, auf ihrem eigenen Gerät abzulegen und zu verwalten. Wenn diese Informationen auch für andere Benutzer von Interesse ist, stellt sich die Frage, wie sie auf diese zugreifen können. Informationen unterschiedlicher Benutzer werden traditionell nur unter menschlicher Einwirkung ausgetauscht. Es ergibt sich für die Benutzer allerdings eine Zeit- und Kostenersparnis, wenn der Austausch von Informationen *automatisiert* durch die Geräte der beteiligten Benutzer erfolgt. Als Voraussetzung dafür muss eine *Systemsoftware*, die den automatisierten Austausch von Informationen ermöglicht, an die menschlichen Benutzer verteilt werden und von ihnen auf ihre jeweiligen Geräte installiert werden. Als Folge davon entsteht ein Informationssystem, dessen Einheiten die Geräte der Benutzer mitsamt der darauf installierten Systemsoftware sind. Dieses Informationssystem ist *selbst-organisierend*, da die Einheiten ohne Koordination einer zentralen Stelle kooperieren. Allerdings entziehen sich die Geräte der menschlichen Benutzer jeder zentralen Kontrolle. Dies ist insofern problematisch, als es menschlichen Benutzern auch ohne Expertenwissen möglich ist, die installierte Systemsoftware und damit die Einheit, die sie im Informationssystem vertritt, zu manipulieren. Aufgrund dieser *Autonomie* der Benutzer kann nicht mehr vorausgesetzt werden, dass sich die Einheiten an die Regeln der Kooperation halten, die im Systementwurf vorgesehen sind. Da die Benutzer unterschiedliche Interessen verfolgen, ist vielmehr zu erwarten, dass sich die Einheiten gegenseitig *betrügen*.

Diese Vision von Informationssystemen, die vollständig auf die Geräte der Endbenutzer verteilt sind, wurde in dieser Arbeit am *Campus-Szenario* festgemacht. Bei diesem handelt es sich um ein Informationssystem, in dem die Geräte von Studenten Informationen automatisiert über ein Ad-hoc-Netz austauschen und sich gegenseitig Dienste erbringen. Aus der Szenariobeschreibung wurde ein Systemmodell abgeleitet, das den weiteren Teilen dieser Arbeit zugrunde lag. Zudem wurden die Techniken der Automatisierung aufgezeigt, die eine Teilnahme der menschlichen Benutzer am Informationssystem im Sinne des Campus-Szenarios ermöglichen. Durch die

Beschreibung des Campus-Szenarios wurde nicht nur anschaulich geklärt, warum die Teilnahme am Informationssystem für menschliche Benutzer von Vorteil ist. Darüber hinaus wurde auch aufgezeigt, *warum* sich *welche* Benutzer zur Manipulation der Systemsoftware entscheiden. Die Verwendung manipulierter Versionen der Systemsoftware führt dazu, dass Betrugsverhalten im Informationssystem vorherrscht. Dies unterminiert die Nützlichkeit des Informationssystems und führt zu Austritten seitens der Benutzer. Letztlich ist das Informationssystem dadurch *nicht existenzfähig*.

Diese Beobachtung wurde bei der Formulierung der *These* dieser Arbeit als Herausforderung ausgelegt: Gesucht ist ein Ansatz, durch den Informationssysteme, die wie im Campus-Szenario vollständig auf die Geräte autonomer Benutzer verteilt sind, existenzfähig sind. Ein solcher Ansatz muss dafür sorgen, dass Betrugsverhalten im Informationssystem nicht möglich ist oder zumindest keine Vorteile mit sich bringt. Letzteres erfordert einen Mechanismus, der betrügende Einheiten bestraft, so dass für sie Betrugskosten anfallen.

Gemäß dem *Stand der Forschung* gibt es traditionell zwei Arten von Ansätzen, die die geforderte Eindämmung von Betrugsverhalten zum Ziel haben: **(1)** Eine Reihe von Ansätzen geht von der Benutzung *manipulationssicherer Hardware* aus, um den menschlichen Benutzern die Kontrolle über ihre Gerät zu entziehen. Dadurch wird von Vornherein verhindert, dass die Systemsoftware manipuliert werden kann. Allerdings haben wir gesehen, dass keine der zwei grundsätzlichen Möglichkeiten zum Einsatz von manipulationssicherer Hardware tauglich ist. Wird die Systemsoftware fest auf ein spezielles Hardware-Modul abgelegt, so entstehen dem Endbenutzer Kosten, die eine zu hohe Eintrittsbarriere für die Teilnahme am Informationssystem darstellen. Mehrzweck-Hardware würde zwar dieses Problem beseitigen, sie ist jedoch nicht verfügbar. **(2)** Eine Gruppe weiterer Ansätze setzt *vertrauenswürdige Dritte* ein, die als zentrale Instanz regulierend in das Informationssystem eingreifen, um Betrugsverhalten unvorteilhaft zu machen. Die Ansätze unterscheiden sich darin, zu welchem Zweck sie Dritte einsetzen. Aufgrund der Annahmen der zugrunde liegenden Austauschprotokolle können die Ansätze zur Konfliktvermeidung und -lösung nur in Spezialfällen eingesetzt werden können. Alternativ dazu sehen einige Ansätze vor, dass der Dritten die Rolle übernimmt, Informationen über das vergangene Verhalten der Einheiten zu sammeln. Dadurch werden betrügende Einheiten mit einer schlechten Reputation gekennzeichnet und zu ihrem Nachteil von weiterer Kooperation ausgeschlossen. Das Hauptproblem der Ansätze mit vertrauenswürdigen Dritten liegt darin, dass die Teilnehmer und der Betreiber des Informationssystems einen enormen Aufwand tragen müssen, damit der Dritte von den Einheiten durchgängig erreichbar ist. Wir haben gesehen, dass somit der Einsatz vertrauenswürdiger Dritte nicht wünschenswert ist.

Weiterhin gibt es Ansätze, die weder den Einsatz manipulationssicherer Hardware noch den von vertrauenswürdigen Dritten bedürfen. Bei ihnen handelt es sich um Ansätze der *verteilten Vertrauensbildung*. Sie bewirken eine gegenseitige Kontrolle der Teilnehmer, indem jede Einheit sich ihren Glauben über die Vertrauenswürdigkeit Anderer bildet. Von der Effektivität einer solchen selbstorganisierenden sozialen Kontrolle hängt es ab, ob Betrugsverhalten im Informationssystem eingedämmt werden kann. Aus dieser Überlegung wurden eine Reihe von Anforderungen identifiziert, anhand derer die Eignung der bestehenden Ansätze zur verteilten Vertrauensbildung bewertet wurde. Hieraus ging hervor, dass die bestehenden Ansätze es verpassen, eine Reihe von Anforderungen zu erfüllen. Das Schlüsselproblem besteht darin, wie die Erfahrungen Anderer in die eigene Glaubensbildung einfließen können. In allen Ansätzen stellen Empfehlungen billiges Gerede dar und erlauben somit keine glaubwürdige Signalisierung eigener Erfahrungen. Die Glaubensbildung kann dadurch nur mit ad-hoc Verfahren durchgeführt werden, die auf Plausibilitätsüberlegungen basieren. Der Glaubensbildung mangelt es daher an notwendiger Präzision

und an Robustheit gegenüber Fehlverhalten. Zudem fehlt es allen Ansätzen an einer Methodik, mit der die eigene Robustheit gegenüber Fehlverhalten überhaupt nachgewiesen werden kann. Wir kamen daher zum Schluss, dass die verteilte Vertrauensbildung nur durch einen grundlegend andersartigen Ansatz zu einer effektiven Betrugsvermeidung führen kann.

11.2 Entwurf

Der eigene Ansatz baut auf die verteilten Vertrauensbildung auf, da sie selbstorganisiernde soziale Kontrolle zwischen autonomen Einheiten ermöglicht. Beim Entwurf des Ansatzes kommt es also darauf an, Erweiterungen und Anpassungen zu finden, die diese soziale Kontrolle und damit die Betrugsvermeidung effektiv machen. Um dies zu erreichen, wurden Schlüsselideen der Ansätze mit manipulationssicherer Hardware und mit Dritten in den eigenen Ansatz aufgenommen: **(1)** Ein Kennzeichen von manipulationssicheren Systemen ist, dass der Systementwurf auch Verhalten vorsehen darf, das individuell nicht rational ist. Der Systementwurf besteht also aus dem Aufstellen von Normen, die die Verhaltensmöglichkeiten einschränken. Ausgehend auf der Beobachtung, dass auch ohne manipulationssichere Hardware der Akt der Manipulation aufwändig ist, kann ein solcher *normativer Systementwurf* auch für den eigenen Ansatz angewendet werden. **(2)** Einige Ansätze mit Dritten setzen nicht-abstreitbare Marken ein, um das Verhalten der Einheiten für den vertrauenswürdigen Dritten nachvollziehbar zu machen. Diese *Beweismittel* unterstützen den Dritten darin, Betrugsverhalten zu erkennen und den davon verursachten Schaden rückgängig zu machen. Um Beweismittel auch im eigenen Ansatz einsetzen zu können, muss sich ihre Anwendung an der Unterstützung der selbstorganisierenden sozialen Kontrolle orientieren.

Bei den *Grundlagen* dieser Arbeiten handelt es sich vor allem um die Konzepte der Spieltheorie. Besondere Bedeutung haben die Erweiterungen der Spieltheorie, die eine asymmetrische Verteilung von Informationen zwischen den Spielern berücksichtigen. Aufgrund dieser Informationsasymmetrie sehen Spieler das Ausführen gewisser Handlungen als Möglichkeit an, ihren Typ anderen Spielern zu signalisieren. Die weiteren Grundlagen umfassen die probabilistische Glaubensrevision unter Unsicherheit und die Problemstellungen der Theorie der verteilten Systeme.

Vor dem eigentlichen Entwurf haben wir uns mit dem Gegensatz zwischen *Systementwurf und Autonomie* befasst und mit den Folgen, die aus ihm entstehen. Jeder menschliche Benutzer entscheidet autonom darüber, ob und mit welcher Software sein Gerät am Informationssystem teilnimmt. Entscheidet er sich für die Verwendung der originalen Systemsoftware (einer manipulierten Version davon), so nennen wir seine Einheit normativ (strategisch). Der Vorteil normativer Einheiten gegenüber strategischer Einheiten ergibt sich daraus, dass der menschliche Benutzer nicht die Hindernisse für die Verwendung manipulierter Versionen der Systemsoftware bewältigen muss. Eine Analyse technischer und rechtlicher Gesichtspunkte hat gezeigt, dass es sowohl für die Erstellung als auch für das Übernehmen einer manipulierten Version Hindernisse gibt, die zu nicht vernachlässigbaren Manipulationskosten führen. Diese sind mit den Normativitätskosten, die bei der Wahl zu einer normativen Einheit auftreten, abzuwägen. Bei ihnen handelt es sich um die Opportunitätskosten, die durch den im Systementwurf vorgesehenen Verzicht auf vorteilhaftes Betrugsverhalten entstehen.

Das eigentliche Ziel des Systementwurfs ist es, die Eigenschaften des Systems zu bestimmen. Jedoch kann der Systementwerfer nicht durchsetzen, dass seine Systemsoftware auch tatsächlich von den autonomen Benutzern verwendet wird. Die Autonomie steht also deswegen im Gegensatz zum Systementwurf, weil sie ihm die direkte Bestimmung der Systemeigenschaften verwehrt. Als Folge davon muss das Ziel des Systementwurfs sein, dass die Benutzer aus ihrem eigenen Interesse

die originale Systemsoftware verwenden. Die Kriterien für die Zielerreichung leiten sich aus denen der Typwahl der Benutzer ab. Um diese Kriterien zu erfüllen, wurden die zwei Maximen des Systementwurfs vorgestellt, die hinreichend kooperatives und hinreichend vorteilhaftes Verhalten für normative Einheiten verlangen. Diese Maximen stehen zwar miteinander im Konflikt, können aber durch die Existenz von Manipulationskosten in Einklang gebracht werden. Eine Umsetzung finden die Maximen in den Verhaltensvorschriften und Normen des Systementwurfs. Die Definition der Vorschriften und Normen orientiert sich maßgeblich daran, dass sie sich im System durchsetzen. Dazu muss ihre Befolgung für die jeweilige Einheit von Vorteil sein.

Als Ausgangspunkt für die Vertrauensbildung macht jede Einheit durch Teilnahme an Transaktionen ihre eigenen Erfahrungen über das Verhalten Anderer. Eine solche *lokale Vertrauensbildung* ist in einem Kreislauf von Transaktionen, Glaubensbildung und Vertrauensentscheidungen organisiert. Die Ablauf von Transaktionen wird zunächst durch das Zwei-Wege Protokoll festgelegt. Den Schwerpunkt bildete die Glaubensbildung. Es wurde gezeigt, dass sie typorientiert sein muss, um die zuvor identifizierten Anforderungen an sie bewältigen zu können. Der Typglaube einer Einheit über eine andere Einheit ist somit als ihre subjektive Wahrscheinlichkeit dafür definiert, dass diese andere Einheit normativ ist. Um eine Beziehung zwischen dem Typ und dem Verhalten einer Einheit herzustellen, wurde das TIB-Modell als Glaubensmodell vorgestellt und es wurde gezeigt, wie dafür benötigte Wahrscheinlichkeit strategischer Einhaltung und unbeabsichtigten Betrugsverhaltens kontextabhängig einzuschätzen ist. Die Vorschriften der Glaubensrevision leiten sich aus dem TIB-Modell ab und sind somit probabilistisch fundiert. Die Odds-Darstellung ermöglicht eine übersichtliche Darstellung der Revisionsvorschriften als Revisionsfaktoren. Zur Bewertung neuer Bekanntschaften wird der Typglauben entsprechend des Systemglaubens initialisiert, dessen Quantifizierung sich nach den bisherigen Erfahrungen einer Einheit richtet. Zum Treffen von Vertrauensentscheidungen wurden Formeln abgeleitet, mit deren Hilfe der erwartete Nutzen einer Transaktionsgelegenheit bewertet werden kann. Diese Bewertung bestimmt über die Teilnahme an einer Transaktion. Erst die Möglichkeit zur Ablehnung von potentiellen Transaktionspartnern erlaubt es den normativen Einheiten, sich trotz des Verzichts auf Betrugsabsichten nicht altruistisch zu verhalten.

Die Erweiterung zur *verteilten Vertrauensbildung* führt zur Verschärfung der sozialen Kontrolle, indem der Austausch der Transaktionserfahrungen der einzelnen Einheiten ermöglicht wird. Die Grundlage hierfür bildet der Einsatz von *transaktionalen Beweismitteln*. Dadurch, dass Festlegungen in Beweismitteln auf nicht-abstreitbare Weise eingegangen werden, entstehen Möglichkeiten zur glaubwürdigen Signalisierung. Der Kreislauf der Vertrauensbildung wurde entsprechend erweitert, um die Ausstellung, Verteilung und Bewertung dieser transaktionalen Beweismittel zu ermöglichen. Zur lokalen Ablage und Verwaltung von Beweismitteln wurde eine Komponente der Beweismittel- und Wissensverwaltung beschrieben und umgesetzt. Wir haben Verträge und Quittungen als die zwei Arten von Beweismitteln identifiziert, deren Ausstellung im Rahmen einer Transaktion sinnvoll ist. Dies führte zu einer Erweiterung des Transaktionsprotokolls zu einem Sechs-Wege Protokoll. Die drei identifizierten Empfehlungsarten stellen allesamt eine Form der indirekten Signalisierung dar. Verträge, Quittungen und negative Empfehlungen bilden jeweils als begleitende transaktionale Beweismittel die Grundlage für negativen Empfehlungen, Selbstempfehlungen und Typbeweise. Eine Sonderrolle nimmt die negative Empfehlung ein, da sich ihr Aussteller in einem entsprechenden Beweismittel auf ihre Aussage festlegt.

Die Berücksichtigung von Empfehlungen führte zu einer Erweiterung der Glaubensbildung. Negative Empfehlungen zeigen einen Konflikt zwischen zwei Einheiten an. Das inkonsistente Verhalten der Einheiten, die ohne Auftreten eines Konflikts dennoch negativ empfehlen, wird mit Hilfe von Selbstempfehlungen und Typbeweisen erkannt. Die entsprechenden Revisionsvorschrif-

ten für die Bewertung von Konflikten und Rehabilitierungen wurden auf probabilistisch fundierte Weise abgeleitet und anschaulich interpretiert. Von entscheidender Wichtigkeit ist, dass bei der Bewertung eines Konflikts diejenige Einheit, die zuerst den Konflikt in einer entsprechenden negativen Empfehlung kundtut, weniger stark abgewertet wird. In der Analyse des Spiels negativer Empfehlungen konnte gezeigt werden, dass dadurch auch strategischen Einheiten zum Empfehlen bereit sind. Somit sichert der Entwurf sowohl die Verfügbarkeit als auch den Wahrheitsgehalt von Empfehlungen. Der Systementwurf ist auch bezüglich der Ausstellung von Verträgen und Quittungen selbstdurchsetzend. Als Folge davon ist das Auftreten eines Konflikts zwischen zwei Transaktionspartnern sowohl eine notwendige als auch eine hinreichende Bedingung für das Ausstellen einer negativen Empfehlung. Da damit die Zahl der Konflikte, an denen eine Einheit beteiligt ist, Aufschluss über ihren Typ gibt, werden die Folgekosten für Betrugsverhalten nachhaltig erhöht und die soziale Kontrolle verschärft.

Normative Einheiten benötigen eine Möglichkeit zum positiven Empfehlen, um sich gezielt von strategischen Einheiten absetzen zu können. Daher wurden die Verfahren der verteilten Vertrauensbildung um den Einsatz *sozialer Beweismittel* erweitert, durch die sowohl aussagekräftige als auch glaubwürdige Selbstempfehlungen ermöglicht werden. In diesen Beweismitteln werden soziale Bindungen zwischen Einheiten festgehalten und das soziale Gefüge somit explizit gemacht. Dadurch wird ermöglicht, dass aus der Stellung einer Einheit im sozialen Gefüge Aufschlüsse über ihren Typ gemacht werden können. Die Untersuchung des Entwurfsraums für soziale Bindungen zeigt, dass die Typ-Bürgschaftsbeziehung diejenige soziale Bindung ist, die einem Informationssystem wie dem des Campus-Szenarios am angemessensten ist. Bei den eingesetzten sozialen Beweismitteln handelt es sich also um nicht-abstreitbare Bürgschaften. Das Empfehlungssystem wurde um die Möglichkeit erweitert, sich durch die Angabe der eigenen Bürgen selbst zu empfehlen.

Um die Kenntnis von Bürgschaftsbeziehungen Anderer in die Glaubensbildung einfließen zu lassen, wurde die Glaubensbildung um die Ebene des sozialen Typglaubens erweitert. Dieser wird nach Bedarf aus dem individuellen Typglauben und den bekannten Bürgschaftsbeziehungen abgeleitet. Der Zusammenhang zwischen individuellem und sozialem Typglauben wird bei der Durchführung von Glaubensrevisionen gewahrt, indem auch der Typglauben über die einzelnen Bürgen revidiert wird. Die Funktionsweise der Vorschriften der Bildung und Revision des sozialen Typglaubens wurde in einem ausführlichen Beispiel dargestellt. Es wurden Vorschriften dafür abgeleitet, wann Einheiten zum Eingehen von Bürgschaftsbeziehungen bereit sind. Die Analyse zeigt, dass nicht nur normative sondern auch strategische Einheiten bevorzugt mit normativen Einheiten Bürgschaftsbeziehungen eingehen. Dies liegt daran, dass das Ausstellen einer Bürgschaft eine Investition darstellt, die sich nur bei kooperativem Verhalten des Gebürgten auszahlt. Schlussendlich wird dadurch erreicht, dass sich normative Einheiten durch die Zahl und der Qualität der eigenen Bürgen glaubhaft von strategischen Einheiten abgrenzen können.

11.3 Evaluation

Der Entwurf des eigenen Ansatzes zielt letztendlich auf die Validierung unserer These, dass die Vision von Informationssystemen, die wie im Campus-Szenario vollständig auf die Geräte autonomer Benutzer verteilt sind, realisiert werden kann. Um diese These validieren zu können, war eine quantitative Bewertung des entworfenen Ansatzes im Rahmen einer Evaluation notwendig.

Für die Evaluation wurde zunächst eine *Methodik* entwickelt. Sie zieht die Simulation als das wesentliche Mittel der Evaluation heran. In diesem Zusammenhang haben wir die entschei-

denden Vorteile der Simulation gegenüber der Analyse aufgezeigt. Aus den Betrachtungen zum Systementwurf wurde ein Kriterium für die Existenzfähigkeit und eine Maßzahl für die Güte des Entwurfs abgeleitet. Vor der eigentlichen Evaluation des Gesamtsystems sind die Möglichkeiten zur Manipulation und dem daraus folgenden Betrugsverhalten auf eine realistische Weise zu berücksichtigen. Ziel dieses Vorschritts ist somit das Auffinden von Gegenstrategien, die als Reaktion manipulationswilliger menschlicher Benutzer auf den normativen Systementwurf zu erwarten sind. Zur computergestützten Durchführung der Evaluation wurden zwei Simulationswerkzeuge vorgestellt. Das Simulative Kooperationsturnier baut auf dem Simulationsrahmenwerk DIANEmu auf, da in ihm das Bewegungsmodell menschlicher Benutzer für das Campus-Szenario eingebaut ist. Dieses Rahmenwerk wurde entsprechend erweitert, um auch andere Aspekte der Rahmenbedingungen des Informationssystems realitätsnah abzubilden. Das Interaktive Kooperationsturnier geht zur Antizipation der Manipulation von der Grundidee aus, dass viel versprechende Gegenstrategien von menschlichen Versuchspersonen im Zuge einer interaktiven Simulation gefunden werden. Voraussetzung hierfür ist die Erlernbarkeit und Bedienbarkeit selbst durch technisch unbedarften Versuchspersonen sowie ihre Motivation zur erfolgreichen Teilnahme. Das Interaktive Kooperationsturnier setzt diese Anforderungen um.

Diese Methodik wurde bei der *Durchführung* der Evaluation eingesetzt. Mit Hilfe des Interaktiven Kooperationsturniers wurden eine Reihe viel versprechende Gegenstrategien gefunden, die in zwei Gruppen eingeteilt werden können. Die einfachen Gegenstrategien zeichnen sich durch ihre kurzfristig ausgerichteten Betrugskriterien aus, während die komplexen Gegenstrategien einen großen Wert auf die initiale Bildung von Vertrauen legen und somit längerfristig angelegt sind. Die gefundenen Gegenstrategien wurden anschließend mit Hilfe des Simulativen Kooperationsturniers bewertet. Aus den Simulationsergebnissen ging hervor, unter welchen Rahmenbedingungen welche Gegenstrategie am erfolgreichsten ist. Das Resultat wurde für den Fall erweitert, dass die manipulationswilligen Benutzer die Rahmenbedingungen des Informationssystems nur unpräzise wahrnehmen können. Durch die Berücksichtigung der jeweils besten Gegenstrategie in der Evaluation des Gesamtsystems wurde eine realitätsnahe und aussagekräftige Simulation des Informationssystems ermöglicht.

Die eigentliche Evaluation des Gesamtsystems wurde entlang einiger Versuchsreihen durchgeführt, die jeweils bestimmte Aspekte der Rahmenbedingungen variieren. Die Simulationsergebnisse validierten unsere These von der Existenzfähigkeit von Informationssystemen, die wie im Campus-Szenario vollständig auf die Geräte der autonomen Teilnehmer verteilt sind. Nur ein sehr geringer Anteil inhärent manipulationsfreudiger Benutzer entscheidet sich zur Verwendung manipulierter Versionen der Systemsoftware. Diese Benutzer können sich aber dadurch weder einen nennenswerten Vorteil verschaffen noch dem Gesamtsystem Schaden zufügen. Dieses Ergebnis haben wir nicht nur für Rahmenbedingungen erhalten, die für das Informationssystem des Campus-Szenarios zu erwarten sind. Die Variation der Modellparameter hat gezeigt, dass für die allermeisten denkbaren Rahmenbedingungen die These von der Existenzfähigkeit des Informationssystems validiert werden kann. Dieses in seiner Klarheit doch überraschende Ergebnis rührt einerseits vom Einsatz der effektiven verteilten Vertrauensbildung des eigenen Ansatzes und andererseits von der Existenz der Hindernisse zur Manipulation. Außerdem konnte dargelegt werden, welche Voraussetzungen erfüllt sein müssen, damit die Existenzfähigkeit des Informationssystems zugesichert werden kann. Grundvoraussetzung ist ein hinreichend hohes Verhältnis zwischen dem Nutzen und den Kosten der Kooperation, damit es zwischen den Einheiten zur initialen Bildung von Vertrauen kommen kann. Ein geringes Problem stellen hingegen die inhärent manipulationsfreudigen und bösartigen Benutzer dar, die sich als einzige Benutzer unter gewissen Bedingungen zur Manipulation entscheiden. Die Evaluation des Gesamtsystems hat zeigen können, dass sogar

viele der bösartigen Benutzer auf die Manipulation verzichten und mit der originalen Systemsoftware am Informationssystem teilnehmen. Damit setzt sich der Systementwurf in besonders nachhaltiger Weise durch.

11.4 Übersicht der Beiträge

Aus der Zusammenfassung der vorangehenden Abschnitte geht hervor, dass in dieser Dissertation eine Reihe von Konzepten und Verfahren vorgestellt wurden. In diesem Abschnitt nehmen wir einen anderen Blickwinkel ein und erörtern, welche Beiträge dieser Arbeit wesentlich sind und den Stand der Forschung nachhaltig voranbringen. Diese Frage wird im Bezug auf die Forschung zur verteilten Vertrauensbildung beantwortet, da der Ansatz dieser Arbeit eben diesem Bereich zuzuordnen ist. Die bestehenden Ansätze dieses Forschungsbereiches wurden bereits in Abschnitt 2.4 vorgestellt. Die wesentlichen Beiträge dieser Arbeit werden im Folgenden mit Blick auf diese Ansätze aufgezeigt.

Problem und Ziel. Was ist das eigentliche Problem von Informationssystemen, die vollständig auf die Geräte autonomer Benutzer verteilt sind, und welches Ziel muss aufgrund dessen ein Ansatz der verteilten Vertrauensbildung verfolgen?

Der Stand der Forschung zu diesen Fragestellungen ist wie folgt: Als Problem wird dargestellt, dass betrügende Einheiten nicht erkannt werden. Ziel ist es daher, eben für diese Erkennung betrügender Einheiten zu sorgen. Dies erklärt, warum alle bestehenden Ansätze, die ihre Effektivität simulativ zu belegen suchen, sich an der Erkennung betrügender Einheiten messen (vergleiche etwa [LI04, DA04]).

Die vorliegende Dissertation zeigt, dass die Identifikation des Problems und Zieles in einen größeren Zusammenhang gebracht werden muss. Konkret sind hierbei die wesentlichen Beiträge:

- Das eigentliche Problem liegt in der Möglichkeit zur Manipulation und der daraus folgenden Degeneration des Informationssystems, die zum Aufhören seiner Existenz führt. Dies wurde anschaulich am Campus-Szenario (Abschnitt 1.2.1) und allgemein an einem Modell (Abschnitt 5.3.2) dargelegt.
- Das Ziel eines Ansatzes der verteilten Vertrauensbildung muss daher sein, die Existenzfähigkeit des Informationssystems zu gewährleisten (Abschnitt 1.3). An diesem Ziel wurde die Evaluation ausgerichtet (Abschnitt 9.1.2).
- Die Durchführung der Evaluation hat zudem gezeigt, dass durch den Einsatz des eigenen Ansatzes dieses Ziel erreicht wird (Abschnitt 10.2). Damit ist klargestellt, dass selbstorganisierende Informationssysteme, deren Einheiten untereinander autonom sind, existenzfähig sind.

Rolle des Systementwurfs. Der Entwurf einer verteilten Vertrauensbildung schlägt sich in der Systemsoftware nieder, die an teilnahmewillige Benutzer verteilt wird. Eine weitere Frage besteht somit darin, inwiefern sich der Entwurf der verteilten Vertrauensbildung im realen Informationssystem wiederspiegelt. Ist der Urheber des Entwurfs ein Systementwerfer oder lediglich jemand, der einen unverbindlichen Vorschlag zum Verhalten der Einheiten unterbreitet?

Bei dieser Frage gehen die Meinungen im Stand der Forschung auseinander. Eine Reihe von Ansätzen erhebt die Anreizkompatibilität als entscheidendes Kriterium für den Entwurf der verteilten Vertrauensbildung [JF03]. Demnach muss es für jede Einheit zu jedem Zeitpunkt von

Vorteil sein, das im Entwurf vorgesehene Verhalten zu zeigen. Ansonsten kommt es zur Manipulation der Systemsoftware [BH03]. Den Nachweis dieser Anreizkompatibilität kann allerdings keiner dieser Ansätze antreten (vergleiche [PS05] und Abschnitt 2.4.3). Eine Gruppe weiterer Ansätze geht davon aus, dass der Anteil der Einheiten, die sich nicht entsprechend des Systementwurfs verhalten, exogen vorgegeben ist (vergleiche [LI04, DA04]). Um diese Einheiten aufzudecken, darf im Systementwurf auch Verhalten vorgeschrieben werden, dass nicht anreizkompatibel ist.

Die Betrachtungen dieser Dissertation haben gezeigt, dass von keinem der bestehenden Ansätze die Rolle des Systementwurfs richtig erkannt wird. In diesem Zusammenhang bringt die vorliegende Arbeit die folgenden wesentlichen Beiträge hervor:

- Es wurde gezeigt, dass die Verwendung manipulierter Versionen der Systemsoftware aufgrund technischer und rechtlicher Hindernisse problematisch ist (Abschnitt 5.2). Dies gilt unabhängig davon, ob ein Benutzer eine manipulierte Version erstellt oder von Anderen übernimmt. Mit der Überwindung dieser technischen und rechtlichen Hindernisse ist ein nicht vernachlässigbarer Aufwand verbunden.

- Es wurden konkrete Faktoren identifiziert, die darüber entscheiden, ob sich ein Benutzer zur Manipulation entscheidet (Abschnitt 5.3.1). Es besteht eine Rückkopplung zwischen dem Verhalten der Benutzer und den Eigenschaften des Informationssystems (Abschnitt 5.1).

- Im Systementwurf muss Verhalten vorgeschrieben werden, das sowohl hinreichend kooperativ als auch hinreichend vorteilhaft ist (Abschnitt 5.4.1). Dieser scheinbare Widerspruch ist nur aufgrund der Existenz der Hindernisse zur Manipulation lösbar. Sie führen zu einem normativen Systementwurf (Abschnitt 5.4.2). Das Kriterium für seine Normen und Vorschriften ist ihre Fähigkeit zur Selbstdurchsetzung (Abschnitt 5.4.3).

Glaubensbildung. Um als Grundlage für utilitaristische Vertrauensentscheidungen herangenommen werden zu können, müssen die Verfahren der Glaubensbildung probabilistisch fundiert sein. Laut Abschnitt 2.4.5 ist dazu unter realistischen Rahmenbedingungen keiner der Ansätze in der Lage. Zudem erlauben die bestehenden Ansätze zur Glaubensbildung keine Aussagen über den Typ anderer Einheiten. In diesem Bereich ist der Beitrag dieser Arbeit wie folgt:

- Das vorgestellte TIB-Modell ermöglicht eine probabilistisch fundierte Glaubensbildung, die sowohl Aussagen über den Typ als auch über das wahrscheinliche Verhalten einer Einheit zulässt (Abschnitt 6.3). Das verwendete Glaubensmodell ist in der Lage, Kontext- und Typinformationen zu berücksichtigen. Die Typorientierung der Glaubensbildung ist die Grundlage für die Verfahren eines normativen Systementwurfs.

Verfügbare, glaubwürdige und aussagekräftige Empfehlungen. Um die soziale Kontrolle zwischen den Einheiten zu verschärfen, bedarf es eines Empfehlungsmechanismus, durch den der Austausch von Transaktionserfahrungen zwischen den einzelnen Einheiten ermöglicht wird. Dieser ist allerdings nur dann nützlich, wenn Empfehlungen tatsächlich ausgestellt werden (also verfügbar sind) und aufgrund des Empfehlungsmodells sowohl glaubwürdig als auch aussagekräftig sind. Abschnitt 2.4.2 hat gezeigt, dass diese Anforderungen in Konflikt stehen.

Laut Abschnitt 2.4 löst keiner der bestehenden Ansätze diesen Konflikt. Entweder werden zusätzliche Anreize für das Ausstellen von Empfehlungen geschaffen, die zur Verminderung der Glaubwürdigkeit von Empfehlungen führen, oder umgekehrt wird das Ausstellen unplausibel erscheinender Empfehlungen bestraft, was die Verfügbarkeit aussagekräftiger Empfehlungen unter-

miniert. Ursache hierfür ist das Empfehlungsmodell, nach dem Empfehlungen beliebig ausgestaltete Glaubensberichte sind. Gemäß dem Stand der Forschung handelt es sich somit um eine offene Fragestellung, der besondere Wichtigkeit zukommt.

In diesem Bereich führt diese Arbeit zwei neuartige Konzepte ein. Die sich ergebenden Beiträge sind wie folgt:

- Die Entwicklung des Konzepts der Beweismittel ist die Grundlage für glaubwürdiges Empfehlen (Abschnitt 7.1.1). Dies rührt daher, dass das Ausstellen und Weitergeben von Beweismitteln einen Mechanismus zur glaubhaften Signalisierung darstellt. Zudem wurden die Prinzipien dargelegt, nach denen Beweismittel im Rahmen der verteilten Vertrauensbildung einzusetzen sind (Abschnitt 7.1.2).

- Transaktionale Beweismittel ermöglichen Empfehlungen über einzelne Transaktionserfahrungen (Kapitel 7). Hierbei wurde gezeigt, welche Erweiterungen sich für den Kreislauf der Vertrauensbildung ergeben. Das entwickelte Empfehlungsmodell unterscheidet sich grundsätzlich von bestehenden Arbeiten und stellt durch seine Basierung auf Beweismitteln die Glaubwürdigkeit von Empfehlungen sicher (Abschnitt 7.3). Die Festlegung der Vorschriften zur Glaubensrevision ist probabilistisch fundiert (Abschnitt 7.4) und gewährleistet die Verfügbarkeit von Empfehlungen (Abschnitt 7.6).

- Die Entwicklung des Konzepts sozialer Bindungen führt in Verbindung mit dem Konzept der Beweismittel dazu, dass sich Einheiten auf eine glaubwürdige und aussagekräftige Weise selbst empfehlen können (Abschnitt 8.1.1 und 8.4.2). Zudem sind diese Selbstempfehlungen ohne Einschränkung verfügbar, da jede Einheit einen inhärenten Anreiz zum Selbstempfehlen besitzt. Darüber hinaus wurde der Entwurfsraum sozialer Bindungen aufgezeigt.

- Das Konzept sozialer Bindungen wurde auf die verteilte Vertrauensbildung angewendet (Abschnitt 8.2). Hierfür wurden Typ-Bürgschaftsbeziehungen als angemessene Form der sozialen Bindung identifiziert (Abschnitt 8.1.1) und die Glaubensbildung um probabilistisch fundierte Verfahren zu ihrer Berücksichtigung erweitert (Abschnitt 8.3).

Evaluationsmethodik. Zwei Fragestellungen sind für die Bewertung eines Entwurfs entscheidend: **(1)** Wie lässt sich bestimmen, ob der Entwurf der verteilten Vertrauensbildung seine Vorgaben erfüllt? **(2)** Unter welchen Bedingungen ist dies der Fall?

Der Stand der Forschung zur simulativen Evaluation ist folgendermaßen: Zur Simulation des Gesamtsystems gibt der Urheber des jeweiligen Ansatzes selbst vor, welche manipulierten Versionen der Systemsoftware zu erwarten sind. Dies führt weder zum systematischen Aufdecken von Angriffspunkten noch zu glaubwürdigen Evaluationsergebnissen. Bei der zweiten Fragestellung beschränken sich die bestehenden Ansätze auf eine Versuchsreihe, die die Zahl der Transaktionsgelegenheiten variiert (vergleiche [LI04, DA04]). Abhängigkeiten von anderen Aspekten der Rahmenbedingungen werden aber nicht berücksichtigt.

Auch in diesem Bereich leistet diese Dissertation durch den Vorschlag einer Evaluationsmethodik einen wesentlichen Beitrag zum Stand der Forschung:

- Viel versprechende Angriffspunkte gegen den Systementwurf werden auf systematische Weise durch das Heranziehen von Versuchspersonen aufgedeckt und in der Simulation des Gesamtsystems berücksichtigt (Abschnitt 9.1.3). Zu diesem Zweck wurde anhand des Interaktiven Kooperationsturniers gezeigt, wie das hierfür benötigte interaktive Simulationswerkzeug zu gestalten ist (Abschnitt 9.3).

- Die Abhängigkeit der Simulationsergebnisse von den angenommenen Rahmenbedingungen wird mit Hilfe einer automatisierten Sensibilitätsanalyse ermittelt (Abschnitt 9.2.2). Die Versuchsreihen aus Abschnitt 10.2 zeigen, wie mit Hilfe eines solchen Werkzeuges die simulative Evaluation durchzuführen ist.

Kapitel 12

君子欲訥于言而敏于行

"Der Edle findet es begehrenswert, den Ruf dafür zu haben, schwerfällig im Reden, aber flink im Handeln zu sein."
(Gespräche und Aussprüche des Konfuzius, 4.24)

Weiterführende Konzepte

Im Hauptteil dieser Arbeit wurde ein Ansatz der verteilten Vertrauensbildung entworfen und evaluiert. Er sorgt dafür, dass Informationssysteme, die wie im Campus-Szenario vollständig auf die Geräte der autonomen Benutzer verteilt sind, existenzfähig sind. In diesem Kapitel werden weiterführende Konzepte vorgestellt, die über den eigenen Ansatz hinausgehen. Dabei werden wir von zwei Fragestellungen geleitet:

- *Erweiterte oder erschwerte Rahmenbedingungen:* Ist der eigene Ansatz auch dann einsetzbar, wenn die Rahmenbedingungen, in denen sich das Informationssystem bewegt, im Vergleich zum Campus-Szenario erweitert oder erschwert sind? Abschnitt 12.1 befasst sich mit der Erweiterung um einseitig vorteilhaften Transaktionen. Verschiedene Arten von Erschwernissen werden in Abschnitt 12.2 besprochen. Um diesen entgegenzutreten, werden weiterführende Konzepte vorgestellt. Sie ergänzen den eigenen Ansatz derart, dass er auch unter diesen erweiterten oder erschwerten Rahmenbedingungen einsetzbar ist.
- *Erweiterungsmöglichkeiten:* Welche Erweiterungen des Entwurfs und der Evaluation der verteilten Vertrauensbildung sind denkbar? Hierzu zeigt Abschnitt 12.3 Konzepte auf, die Aspekte des Entwurfs und der Evaluation sinnvoll erweitern.

Dieses Kapitel geht über das Aufzeigen zukünftiger Forschungsrichtungen hinaus, indem es für jede der aufgeworfenen Fragestellungen Konzepte zu deren Lösung vorstellt. In diesem Punkt unterscheidet sich dieses Kapitel vom Ausblick, der im nachfolgenden Kapitel 13 gegeben wird. Dieser begnügt sich damit, relevante Fragestellungen zu identifizieren, die von zukünftigen Forschungsarbeiten zu bearbeiten sind.

12.1 Kooperationsanreize bei einseitig vorteilhaften Transaktionen

Bisher sind wir von Transaktionen ausgegangen, die für die Transaktionspartner von beidseitigem Vorteil sind. In diesem Abschnitt gehen wir auf die Frage ein, welche Änderungen sich

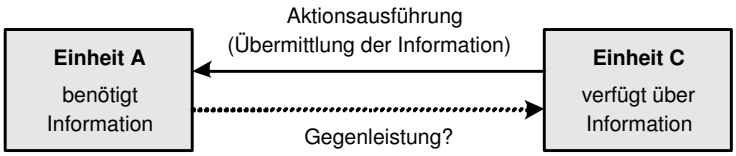

Abbildung 12.1: Die Ausgangslage einer einseitig vorteilhaften Transaktion

ergeben, wenn dem nicht der Fall ist. Hierfür stellen wir in Abschnitt 12.1.1 zunächst Vorüberlegungen zu dieser Fragestellung an. Es zeigt sich, dass das Zustandekommen einseitig vorteilhafter Transaktionen notwendig ist, um Verzögerungen beim Erhalt benötigter Informationen zu vermeiden. Um dies zu erreichen, wird ein prinzipieller Ansatz aufgezeigt, der auf Versprechen basiert. Auf diesem Ansatz aufbauend werden weiterführende Konzepte vorgestellt, die die Versprechen unterschiedlich ausgestalten. Ausgangspunkt hierfür ist das Konzept der Eigenwechsel, das in Abschnitt 12.1.2 besprochen wird. Eine Erweiterung dieses Konzepts um Inhaberwechsel wird in Abschnitt 12.1.3 aufgezeigt. Abschließend gibt Abschnitt 12.1.4 eine Übersicht der Anreize, die eine Einheit zur Teilnahme an einer Transaktion bewegen können.

12.1.1 Versprechen als Ersatz für fehlende Gegenleistungen

Im Folgenden stellen wir grundsätzliche Überlegungen dazu an, warum das Zustandekommen einseitig vorteilhafter Transaktionen wünschenswert ist. Anschließend wird der Einsatz von Versprechen als ein viel versprechender Ansatz identifiziert, der dieses Zustandekommen ermöglicht.

Notwendigkeit einseitig vorteilhafter Transaktionen. Anhand des Campus-Szenarios aus Abschnitt 1.2.1 lässt sich aufzeigen, warum einseitig vorteilhafte Transaktionen sinnvoll sind. Gemäß der Szenariobeschreibung tauscht Anna ihren Vorlesungsmitschrieb gegen Claudes Mensaplan. Diese Transaktion ist von beidseitigem Vorteil, da beide von ihrem jeweiligen Gegenüber Informationen erhalten, die von Interesse sind. Wie ändert sich jedoch die Situation, wenn Anna auf den Mensaplan zugreifen will, bevor sie den Vorlesungsmitschrieb erstellt hat? In diesem Fall könnte Claudes Gerät zwar weiterhin den Mensaplan zur Verfügung stellen. Die Transaktion wäre aber nur noch für Anna von Vorteil, da Claude keine Gegenleistung erhält. Nach unserem bisherigen Verständnis würde Claudes Gerät daher erst in eine Transaktion einwilligen, wenn Anna ihren Vorlesungsmitschrieb abgeschlossen hat und ihn als Gegenleistung zur Verfügung stellen kann. Dadurch ergibt sich aus Sicht Annas eine Verzögerung für den Zugriff auf den benötigten Mensaplan. Es stellt sich damit die Frage, ob diese Verzögerung zwingend ist.

Abbildung 12.1 stellt diese Ausgangslage schematisch dar: Einheit A (Annas Einheit) ist an einer Information (dem Mensaplan) interessiert, über die eine Einheit C (Claudes Einheit) verfügt. Einheit A kann allerdings zu diesem Augenblick keine Information anbieten, die für Einheit C von Interesse ist (der Vorlesungsmitschrieb ist noch nicht erstellt). Eine Transaktion, in der Einheit C die benötigte Information preisgibt, wäre somit nur für Einheit A von Vorteil. Sie kann also nur dann stattfinden, wenn Einheit C einen entsprechenden Anreiz von Einheit A erhält. Der Erhalt der benötigten Information verzögert sich für Einheit A so lange, bis sie einen solchen Anreiz geben kann. Um diese Verzögerung zu eliminieren, sind weiterführende Konzepte gefragt, mit Hilfe derer derartige Anreize gegeben werden können.

Versprechen als Anreiz. Welche Art von Anreiz kommt in Frage? Da Einheit C im Interesse seines menschlichen Prinzipals handelt, zahlt sich für sie die Preisgabe ihrer Information nur dann aus, wenn auch sie in Zukunft Informationen von Einheit A benötigen wird. Als Anreiz bietet sich somit an, dass Einheit A *verspricht*, sich zu einem späteren Zeitpunkt Einheit C gegenüber erkenntlich zu zeigen. Im Folgenden wird basierend auf dieser Überlegung erörtert, wie solche Versprechen Eingang in das Informationssystem und die verteilte Vertrauensbildung finden können.

Der Einsatz von Versprechen macht nur dann Sinn, wenn der Adressat des Versprechens einen echten Anreiz erhält, die benötigte Information zur Verfügung zu stellen. Warum sollte also Einheit C an der für sie unvorteilhaften Transaktion teilnehmen, wenn Einheit A ihr eine zukünftige Gegenleistung verspricht? Weigert sich Einheit C zur Teilnahme an der Transaktion, so erhält sie kein solches Versprechen. Umgekehrt macht also die Teilnahme nur dann Sinn, wenn das Versprechen aus der Sicht von Einheit C einen Wert für sich darstellt. Auf den ersten Blick erscheint es unmöglich, dass einem Versprechen, dessen Einhaltung aufgrund der Autonomie der Einheiten nicht erzwungen werden kann, ein Wert zukommt. Das Bild ändert sich allerdings, wenn nicht nur der Adressat sondern auch der Urheber des Versprechens bei seiner Nichteinhaltung einen Schaden erleidet. Konkret bedeutet dies, dass Einheit A Kosten tragen muss, wenn sie ihr Versprechen einer zukünftigen Gegenleistung nicht einhält.

Wie lässt sich erreichen, dass die Nichteinhaltung von Versprechen Folgekosten verursacht? Die Antwort auf diese Frage ergibt sich analog zu den Überlegungen darüber, wie transaktionales Betrugsverhalten eingedämmt werden kann. Dieses Betrugsverhalten führte unter Einsatz der verteilten Vertrauensbildung zu Folgekosten, da aufgrund der sozialen Kontrolle die betrügende Einheit seltener als Transaktionspartner angenommen wird und ihr somit die Vorteile aus der Teilnahme an Transaktionen entgehen. Damit die Nichteinhaltung von Versprechen mit Folgekosten verbunden ist, müssen wir also die Verfahren der verteilten Vertrauensbildung auf das Verhalten ausdehnen, das die Einheiten bezüglich ihrer Versprechen zeigen. Konkret bedeutet dies Folgendes: Nur einer Einheit, die ihre bisherigen Versprechen eingehalten hat, wird geglaubt, dass sie auch ihre zukünftigen Versprechen einlösen wird. Das Versprechen einer Einheit wird somit nur dann als wertvoll erachtet, wenn aufgrund ihres früheren Verhaltens darauf vertraut werden kann, dass dieses Versprechen auch tatsächlich eingehalten werden wird.

Zusammenfassend lässt sich somit sagen, dass eine Einheit deswegen ihre Versprechen einhält, damit ihre Versprechen auch in Zukunft angenommen werden. Nur so kann sie andere Einheiten dazu bewegen, in einseitig vorteilhafte Transaktionen teilzunehmen.

Abstreitbare Versprechen. Welche Auswirkungen hat es, wenn Versprechen abstreitbar sind? In diesem Fall wissen von einem Versprechen nur die Einheit, die das Versprechen gibt, und die, die das Versprechen erhält. Unbeteiligte Einheiten können somit nicht glaubhaft vom Zustandekommen des Versprechens oder seiner Einlösung unterrichtet werden.

Dies hat direkte Auswirkungen auf die Vertrauensbildung: Hält eine Einheit ihr Versprechen nicht ein, so wird sie nur von der Einheit abgewertet, der sie das Versprechen gegeben hat. Die betrogene Einheit besitzt im Rahmen des beweismittelbasierten Empfehlungssystem keinerlei Möglichkeit, den Betrüger negativ zu empfehlen. Dadurch, dass Verhalten bezüglich des Gebens und Einhaltens von Versprechen nicht in Empfehlungen eingeht, ist die Vertrauensbildung im Bezug auf Versprechen somit (im Sinne von Kapitel 6) lediglich lokal.

Diese Überlegungen zeigen, dass die Folgekosten für die Nichterfüllung eines abstreitbaren Versprechens beschränkt sind. In den folgenden Abschnitten werden weiterführende Konzepte aufgezeigt, die durch eine entsprechende Ausgestaltung von Versprechen dieses Problem beseiti-

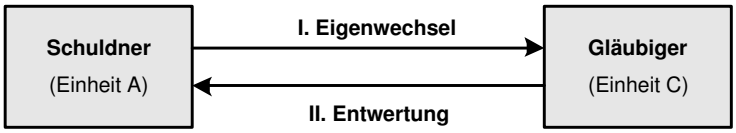

Abbildung 12.2: Rollen und Beweismittelarten beim Einsatz von Eigenwechseln

gen.

12.1.2 Konzept der Eigenwechsel

Im Folgenden wird ein Konzept eingeführt, das auf die Nichtabstreitbarkeit von Versprechen beruht. Hierzu untersuchen wir zunächst, wie nicht-abstreitbare Versprechen (so genannte Eigenwechsel) geartet sein müssen, um die verteilte Vertrauensbildung effektiv zu unterstützen. Anschließend erörtern wir den Einsatz der Eigenwechsel im Empfehlungssystem und in der Glaubensbildung. Dieser Abschnitt schließt mit einer Diskussion der offenen Fragestellungen, die für den Einsatz von Eigenwechseln in Zukunft zu bearbeiten sind.

Eigenwechsel als nicht-abstreitbare Versprechen. Die Ausgestaltung von Versprechen muss sich daran orientieren, dass sie effektiv in der verteilten Vertrauensbildung eingesetzt werden können. Dazu ist erforderlich, dass auch unbeteiligte Einheiten in der Lage sind nachzuvollziehen, welche Versprechen gegeben und eingehalten worden sind. Zu diesem Zweck müssen Versprechen nicht-abstreitbar sein. Das Konzept der Beweismittel lässt sich somit auf die bisherigen Überlegungen anwenden. Das Ergebnis ist das Konzept der *Eigenwechsel* (vergleiche [ON03]): Ein Eigenwechsel ist ein Beweismittel, dessen Aussage darin besteht, dass eine zukünftige Gegenleistung versprochen wird. Im Ausgangsbeispiel muss also Einheit A einen Eigenwechsel ausstellen und an Einheit C übermitteln. Bei einem solchen Eigenwechsel nimmt Einheit A die Rolle des *Schuldners* und Einheit C die des *Gläubigers* ein. Kommt es später zur Erbringung der im Eigenwechsel versprochenen Gegenleistung, so ist es an Einheit C, eine *Entwertung* an Einheit A zu übermitteln. Dabei handelt es sich um eine weitere Art von Beweismittel. Die Aussage einer Entwertung besteht darin, dass das Versprechen, das in einem Eigenwechsel gegeben worden ist, eingehalten wurde und der Eigenwechsel somit entwertet ist. Der Zusammenhang zwischen den Rollen und Beweismittelarten, die sich aus dem Konzept der Eigenwechsel ergibt, wird in Abbildung 12.2 resümiert.

Abbildung 12.3 zeigt den zeitlichen Ablauf, der dem Einsatz von Eigenwechseln und Entwertungen zugrunde liegt. In einer *initialen Transaktion* führt Einheit C die von Einheit A benötigte Aktion_C aus und erhält hierfür den Eigenwechsel, den Einheit A ausstellt. Der Ablauf dieser Transaktion richtet sich nach dem Sechs-Wege Protokoll aus Abschnitt 7.2.2. Der einzige Unterschied liegt darin, dass im Rahmen der initialen Transaktion die Aktionsausführung der Einheit A im Ausstellen und Übermitteln des Eigenwechsels besteht. Das Einlösen ihres Versprechens erfolgt in einer weiteren Transaktion, der *Einlösungstransaktion*. Hierbei führt Einheit C zwar keine Aktion im eigentlichen Sinne für Einheit A aus. Sie entwertet jedoch den zuvor erhaltenen Eigenwechsel durch das Ausstellen einer entsprechenden Entwertung. Im Gegenzug führt Einheit A die versprochene Aktion_A aus. Der Einsatz des Sechs-Wege Protokolls für diese beiden Transaktionen ist sinnvoll, da der zusätzliche Austausch von Verträgen und Quittungen vor Betrugsverhalten in einer der beiden Transaktion abschreckt. Zum Beispiel könnte Einheit A in der Einlösungstransaktion die Ausführung der Aktion_A verweigern, nachdem ihr Eigenwechsel bereits von Einheit C

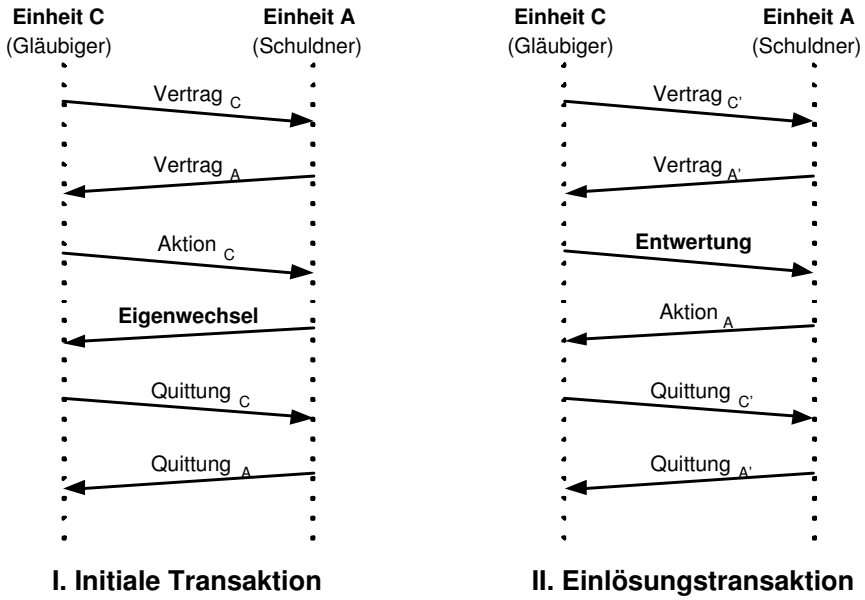

Abbildung 12.3: Ausstellung von Eigenwechseln und Entwertungen in Transaktionen

entwertet worden ist. In diesem Fall wäre die betrogene Einheit C jedoch im Besitz von Vertrag$_{A'}$, den sie zum Ausstellen einer entsprechenden negativen Empfehlung benutzen wird.

Eigenwechsel und die verteilte Vertrauensbildung. Den Kern des Konzepts der Eigenwechsel bildet die Möglichkeit zur glaubhaften Signalisierung: Durch das Ausstellen eines Eigenwechsels signalisiert der Schuldiger seinem Gläubiger, dass er gedenkt, sein Versprechen einzulösen. Damit diese Signalisierung glaubhaft ist, müssen Eigenwechsel und Entwertungen Eingang in die verteilte Vertrauensbildung finden. Mit dieser Aufgabe beschäftigen wir uns im Folgenden.

Als Voraussetzung für den Einsatz von Eigenwechseln und Entwertungen muss jede Einheit in der Lage sein, diese Arten von Beweismitteln bei sich lokal zu verwalten und anderen Einheiten in Empfehlungen weiterzugeben. Die Fähigkeit zur lokalen Verwaltung wird durch eine entsprechende Erweiterung des Regelsystems der Beweismittel- und Wissensverwaltung gewährleistet. Abschnitt A.1.1 des Anhangs zeigt, wie dies erfolgen muss. Zur Einbeziehung von Eigenwechseln und Entwertungen in das Empfehlungssystem müssen wir uns vor Augen führen, welcher Zusammenhang zwischen diesen beiden Arten von Beweismitteln und den transaktionalen Beweismitteln aus Kapitel 7 besteht. Der einzige Unterschied zwischen Vertrag und Eigenwechsel liegt darin, dass das Versprechen eines Vertrags in derselben Transaktion eingelöst wird, während dies beim Eigenwechsel in einer separaten Einlösungstransaktion erfolgt. Entwertungen und Quittungen stehen in einem analogen Zusammenhang. Für das Empfehlungssystem ergibt sich somit folgende Erweiterung:

- *Negative Empfehlungen:* Der Gläubiger eines Eigenwechsels kann den Schuldner anderen Einheiten gegenüber negativ empfehlen, indem er dessen Eigenwechsel in die negative Empfehlung einbindet.

- *Selbstempfehlungen:* Eine Einheit, die das Versprechen ihres Eigenwechsels (das so genannte

Wechselversprechen) eingelöst hat, kann sich anderen Einheiten gegenüber selbst empfehlen, indem sie die vom Gläubiger ausgestellte Entwertung des Eigenwechsels vorzeigt.

- *Typbeweise:* Hat ein Gläubiger seinen Schuldner negativ empfohlen, obwohl er dessen Eigenwechsel entwertet hat, so hat er sich inkonsistent verhalten. In diesem Fall lässt sich durch Vorlage der Entwertung und der negativen Empfehlung beweisen, dass der Gläubiger strategisch ist.

Im Empfehlungssystem werden Eigenwechsel und Entwertungen also gleich wie Verträge und Quittungen behandelt. Somit ist keine Erweiterung des Empfehlungssystems erforderlich. Eine direkte Folge davon ist, dass keine zusätzlichen Vorschriften zur Glaubensrevision benötigt werden: Wenn ein Gläubiger seinen Schuldner negativ empfiehlt, lässt sich ableiten, dass entweder sich der Schuldner weigert, sein Versprechen einzulösen, oder der Gläubiger den Eigenwechsel nicht entwerten will. Folglich hat entweder der Gläubiger oder der Schuldner betrogen. Dies entspricht der Definition eines Konflikts zwischen zwei Einheiten aus Abschnitt 7.4.1. Somit können die uns bekannten Revisionsvorschriften zur Bewertung von Konflikten angewendet werden.

Offene Fragestellungen. Die Ausführungen haben zwar gezeigt, dass das Konzept der Eigenwechsel einen viel versprechenden Ansatz dazu darstellt, dass einseitig vorteilhafte Transaktionen zustande kommen. Allerdings existieren eine Reihe offener Fragestellungen, die sich aus dem Einsatz dieses Konzepts ergeben. Mit diesen beschäftigen wir uns im Folgenden.

Die *Bestimmung der Wechselversprechen* ist keineswegs trivial. Zum Zeitpunkt der Ausstellung eines Eigenwechsels ist unter Umständen nicht klar, welche Aktion in Zukunft vom Schuldiger ausgeführt werden kann und vom Gläubiger benötigt wird. Das Versprechen, das im Eigenwechsel gemacht wird, kann sich also nicht immer auf eine konkrete Aktion beziehen. Dasselbe Problem ergibt sich bei der Festlegung des Zeitraums, in dem das Wechselversprechen eingelöst werden muss. Ohne Bezug auf eine konkrete zukünftige Aktion ist ungewiss, wie dieser Zeitraum zu bestimmen ist.

Ein weiteres offenes Problem stellt die *Bewertung des Einlöseverhaltens* dar: Wenn der Schuldner sein Wechselversprechen einhält, so verhält er sich aus der Sicht des Gläubigers kooperativ. Dies führt mit Hilfe der uns bekannten Verfahren dazu, dass der Gläubiger seinen Typglauben über den Schuldner revidiert. Wie ist allerdings der Fall zu bewerten, dass der Schuldner sein Wechselversprechen bricht? Es gibt vielfältige Gründe dafür, dass sich dies auch ohne Betrugsabsicht seitens des Schuldners ereignet. Er könnte unerwarteterweise nicht in Besitz der Information gekommen sein, deren Übermittlung er dem Gläubiger versprochen hat. Eine weitere Möglichkeit besteht darin, dass der Schuldner nicht mit dem Gläubiger kommunizieren kann, da sich ihre menschlichen Prinzipale zu weit voneinander entfernt haben. Diese Überlegungen zeigen zweierlei: Einerseits ist die Wahrscheinlichkeit für das unbeabsichtigte Brechen von Wechselversprechen weitaus größer als die für unbeabsichtigten Betrug in einfachen Transaktionen. Andererseits ist diese Wahrscheinlichkeit sehr schwer einzuschätzen, da es außer den Kommunikationsabbrüchen eine Reihe weiterer möglicher Ursachen für das unbeabsichtigte Brechen von Wechselversprechen gibt.

Wie entscheidet eine Einheit, ob sie bereit ist, die Rolle des Gläubigers einzunehmen? Diese Entscheidung ist abhängig vom Glauben der Einheit und stellt somit eine Vertrauensentscheidung dar. Damit sie utilitaristisch getroffen werden kann, muss der *Wert des Eigenwechsels* eingeschätzt werden können. Zu diesem Zweck ist die Wahrscheinlichkeit dafür zu bestimmen, dass es auch tatsächlich zum Einlösen des Wechselversprechens kommt. Hierzu reicht es nicht aus, dass die Wahrscheinlichkeit für unbeabsichtigtes Brechen von Wechselversprechen trotz der oben

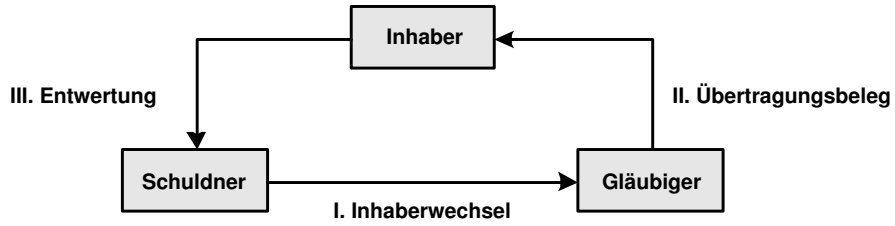

Abbildung 12.4: Rollen und Beweismittelarten beim Einsatz von Inhaberwechseln

genannten Schwierigkeiten abgeschätzt werden kann. Zudem bedarf es einer Einschätzung, wie wahrscheinlich das Einlösen des Versprechens durch einen strategischen Schuldner ist. Dabei ist das Wissen darüber einzubeziehen, wie viele Eigenwechsel des Schuldners noch nicht eingelöst worden sind.

Ein weiteres Problem ergibt sich für das Empfehlungssystem: Der Gläubiger ist zwar in der Lage, den Schuldner auf der Basis dessen Eigenwechsels negativ zu empfehlen. Allerdings verfügt auf der anderen Seite der Schuldner über keinerlei Beweismittel, mit dem er den Gläubiger negativ empfehlen könnte. Dies ist insofern ein Problem, als der komparative Vorteil des Erstempfehlers im Sinne von Abschnitt 7.6.2 wegfällt. Somit besitzt der Gläubiger keinen *Anreiz zum negativen Empfehlen*. Um dieses Problem zu lösen, müsste der Gläubiger initial in einem Beweismittel dem Schuldner bestätigen, dass er die Rolle des Gläubigers einnimmt. Mit diesem Beweismittel könnte auch der Schuldner negativ empfehlen, so dass der komparative Vorteil des Erstempfehlens aufrechterhalten werden könnte.

Die aufgeworfenen Fragestellungen zeigen, dass für die Umsetzung des Konzepts der Eigenwechsel weitere Forschungsarbeiten vonnöten sind. Die Betrachtungen dieses Abschnitts können diesen Arbeiten als Ausgangspunkt dienen.

12.1.3 Konzept der Inhaberwechsel

Eigenwechsel stellen keineswegs die einzige Möglichkeit dar, zukünftige Gegenleistungen zu versprechen. Im Folgenden werden die so genannten Inhaberwechsel als eine Alternative zu den Eigenwechseln aufgezeigt. Dies führt zu einer Erweiterung des Konzepts der Eigenwechsel.

Inhaberwechsel als übertragbare Eigenwechsel. Das Konzept der Eigenwechsel lässt sich dadurch erweitern, dass die direkte Bindung des Wechselversprechens an den Gläubiger aufgelöst wird. Dies bedeutet, dass der Gläubiger das Anrecht, die Einlösung des Wechselversprechens beim Schuldner einzufordern, einer anderen Einheit überlassen kann. Die Einheit, die dieses Anrecht besitzt, nennen wir den *Inhaber* des Wechsels. Den Wechsel selbst bezeichnen wir folgerichtig als *Inhaberwechsel*. Überträgt der Inhaber eines Inhaberwechsels das Anrecht auf dessen Einlösung einer anderen Einheit, so muss er ihr hierfür einen nicht-abstreitbaren *Übertragungsbeleg* ausstellen. Bei diesem handelt es sich ebenfalls um ein Beweismittel. Die Rollen und Beweismittelarten, die bei Inhaberwechseln auftreten, sind in Abbildung 12.4 dargestellt. Aus dem Vergleich mit Abbildung 12.2 wird deutlich, dass das Konzept der Inhaberwechsel dasjenige der Eigenwechsel verallgemeinert.

Warum fordern wir, dass der Übertragungsbeleg nicht-abstreitbar ist? Diese Nichtabstreitbarkeit bringt zwei Vorteile mit sich: Zum einen ist der Inhaber in der Lage, auch in Abwesenheit des Gläubigers den Schuldner davon zu überzeugen, dass er der Begünstigte der versprochenen

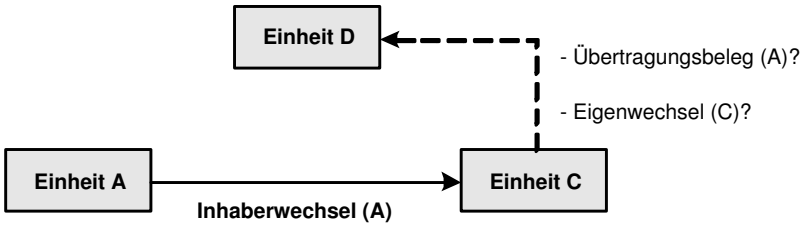

Abbildung 12.5: Beispiel für den Vorteil von Inhaberwechseln gegenüber Eigenwechseln

Gegenleistung ist. Zum anderen verbietet es sich für den Gläubiger, das Anrecht auf die Einlösung desselben Wechselversprechens auf mehrere Einheiten zu übertragen. Da bei jeder Übertragung das Ausstellen eines entsprechenden Beweismittels anfällt, wären nämlich andere Einheiten in der Lage, solches inkonsistentes Verhalten zu erkennen. In der Folge würde der Gläubiger als strategisch angesehen werden.

Vorteile gegenüber Eigenwechseln. Inhaberwechsel besitzen einige Vorteile gegenüber Eigenwechseln. Auf diese gehen wir im Folgenden ein.

Durch den Einsatz von Inhaberwechseln wird die *Zahl der Wechselversprechen verringert*. Dies wird offensichtlich, wenn wir eine Erweiterung des Ausgangsbeispiels betrachten, die in Abbildung 12.5 dargestellt ist. Initial übergibt Einheit A an Einheit C einen Inhaberwechsel. Zu einem späteren Zeitpunkt ist Einheit C an einer Information interessiert, über die eine weitere Einheit D verfügt. Allerdings besitzt Einheit C keine Information, die für Einheit D von Interesse wäre. Nach dem bisherigen Vorgehen würden wir in dieser Situation fordern, dass auch Einheit C der Einheit D einen Eigenwechsel ausstellt. Durch den Einsatz von Inhaberwechseln gibt es allerdings eine Alternative hierzu. Einheit C könnte der Einheit D den Inhaberwechsel übertragen, den sie initial von Einheit A erhalten hat. Somit entfällt für Einheit C der Schritt, ein eigenes Wechselversprechen zu geben. Zusammenfassend sind also durch den Einsatz von Inhaberwechseln weniger Wechselversprechen notwendig.

Ein weiterer Vorteil ergibt sich aus der *größeren Akzeptanz* der Inhaberwechsel. Eigenwechsel werden nur von den Einheiten angenommen, die als hinreichend vertrauenswürdig erscheinen. Somit können die Einheiten nur sehr bedingt von Eigenwechseln Gebrauch machen. Im vorigen Beispiel würde etwa Einheit D unter Umständen keinen Eigenwechsel von Einheit C annehmen wollen, da sie noch keine Erfahrungen mit Einheit C gemacht hat. Dieses Problem ergibt sich nicht, wenn Inhaberwechsel zum Einsatz kommen. In diesem Fall reicht es aus, initial von einer allgemein als vertrauenswürdig akzeptierten Einheit einen Inhaberwechsel zu erhalten. Wenn im Beispiel etwa Einheit D Einheit A als vertrauenswürdig erachtet, so würde sie von Einheit C die Übertragung des Inhaberwechsels als Gegenleistung akzeptieren. Wie lässt sich erklären, dass eine Übertragung des Inhaberwechsels eher als ein neuer Eigenwechsel angenommen wird? Der Unterschied liegt darin, dass Betrug bei der Übertragung von Inhaberwechseln mit hohen Folgekosten verbunden ist. Sie entstehen dadurch, dass die betrügende Einheit durch entsprechende Typbeweise von anderen Einheiten als strategisch erkannt wird. Im Gegensatz dazu führt die Nichteinhaltung eines Eigenwechsels lediglich zu einer negativen Empfehlung. Zusammenfassend lässt sich somit feststellen, dass bei der Verwendung von Inhaberwechseln auch nicht etablierte Einheiten von Wechselversprechen Gebrauch machen können.

Der dritte Vorteil betrifft die *vereinfachte Bestimmung von Wechselversprechen*. Bei Inhaberwechseln ist der Inhaber der Begünstigte des Wechselversprechens. Sollte also eine Einheit

im Besitz eines Inhaberwechsels sein, dessen Wechselversprechen für sie uninteressant geworden ist, so kann sie den Inhaberwechsel an eine andere Einheit übertragen. Diese Flexibilität ermöglicht einen größeren Spielraum für die initiale Bestimmung des Wechselversprechens. Ein Beispiel hierfür lässt sich anhand des Campus-Szenarios aufzeigen: Bobs Einheit könnte Annas Einheit in einem Inhaberwechsel versprechen, in einer zukünftigen Transaktion gewisse Dateiformate zu konvertieren. Dabei ist es für Annas Einheit unerheblich, ob sie in Zukunft eine solche Konversion tatsächlich benötigen wird. Im Zweifelsfall kann sie dieses Wechselversprechen an andere Einheiten übertragen. Dies ist insbesondere dann von Vorteil, wenn sich Annas und Bobs Gerät über einen längeren Zeitraum hinweg nicht in Kommunikationsreichweite befinden. Dann muss Anna das Wechselversprechen nicht verfallen lassen, wie es etwa beim Eigenwechsel der Fall wäre, sondern sie kann ihn anderen Einheiten übertragen.

Die Besprechung der Vorteile des Konzepts der Inhaberwechsel zeigt, dass sein Einsatz zu grundlegenden Änderungen im Informationssystem führt. Die Inhaberwechsel von Einheiten, die allgemein als vertrauenswürdig erachtet werden, sind begehrt, da sie von vielen Einheiten als Gegenleistung akzeptiert werden. Als Folge davon kommt es unter Umständen dazu, dass die Inhaberwechsel einiger weniger Einheiten als de facto Währung im System kursieren. Damit verbunden ist die Flexibilität bei der Bestimmung von Wechselversprechen. Damit Inhaberwechsel ihre Stellung als Währung wahrnehmen können, müssen sich diese Wechselversprechen auf standardisierte Gegenleistungen beziehen, die allen Einheiten mehr oder weniger benötigen. Damit wird sichergestellt, dass jede Einheit begierig darauf ist, Inhaberwechsel übertragen zu bekommen. Insgesamt ergibt sich also durch die Erweiterung zu Inhaberwechseln eine Reihe grundlegender Änderungen im Gesamtsystem. Wie diese zu bewerten und zu behandeln sind, ist eine offene Fragestellung, die in zukünftigen Forschungsarbeiten angegangen werden muss.

Nachteile bei der Umsetzung. Auf der anderen Seite ergeben sich durch die Verwendung von Inhaberwechseln auch Nachteile. Sie rühren vom Aufwand, den die Übertragung der Anrechte auf das Wechselversprechen mit sich bringt. Dies führt nicht nur zu einem erhöhten Aufwand im Zuge der Übertragung. Darüber hinaus wird der Einsatz von Inhaberwechseln im Empfehlungssystem erschwert. Zum Ausstellen einer Selbstempfehlung reicht es nicht mehr aus, eine Entwertung des ausgestellten Wechsels vorzuweisen. Darüber hinaus müssen die Übertragungsbelege ausgewiesen sein, die dem Aussteller der Entwertung die Vollmacht gaben, den Inhaberwechsel zu entwerten. Entsprechendes gilt für das negative Empfehlen.

Die Notwendigkeit dieses Zusatzaufwands wird deutlich, wenn wir uns einer Erweiterung des Ausgangsbeispiels zuwenden, die in Abbildung 12.6 dargestellt ist: Einheit C überträgt den Inhaberwechsel der Einheit A an Einheit D. Anschließend kommt es zur Einlösungstransaktion zwischen Einheit A und D, in der der Inhaberwechsel von Einheit D entwertet wird. Wie kann sich Einheit A verteidigen, wenn Einheit C sie einer anderen Einheit E gegenüber negativ empfiehlt? Einheit C ist zu einer solchen negativen Empfehlung in der Lage, da sie bei der Übertragung des Inhaberwechsels keineswegs gezwungen ist, ihn bei sich lokal zu löschen. Um die negative Empfehlung zu widerlegen, reicht die von Einheit D ausgestellte Entwertung nicht aus. Diese beweist nämlich nicht, dass Einheit D überhaupt das Anrecht auf das Wechselversprechen hatte. Zudem muss also auch der Übertragungsbeleg vorgezeigt werden. Ein erhöhter Aufwand ergibt sich auch für Einheit D, wenn sie Einheit A negativ empfehlen will. Hierfür muss sie aus analogen Gründen außer dem Inhaberwechsel auch ihren Übertragungsbeleg vorzeigen.

Bewertung. Der Einsatz von Inhaberwechseln führt zwar zu einem gewissen Aufwand und zu einer erhöhten Komplexität des Entwurfs. Auf der anderen Seite besitzen aber Inhaberwechsel eine

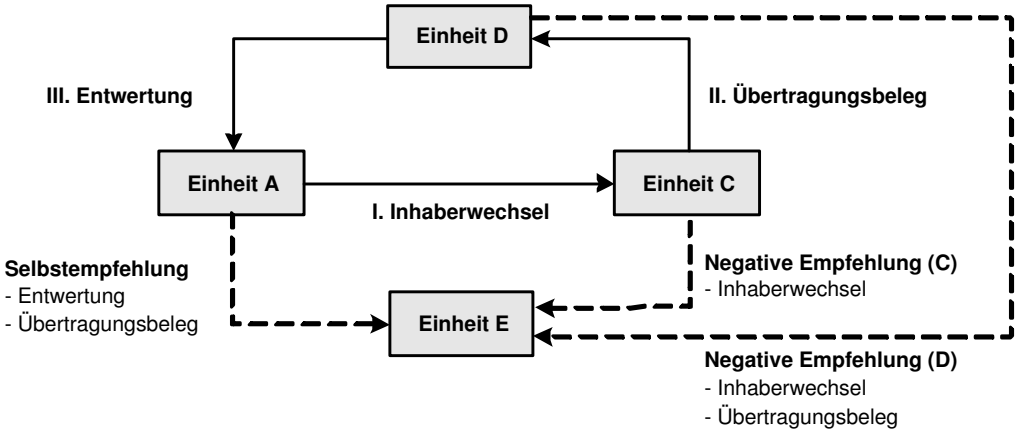

Abbildung 12.6: Beispiel für das Zusammenspiel zwischen Inhaberwechseln und Empfehlungen

Reihe von wesentlichen Vorteilen gegenüber Eigenwechseln. Die Verwendung von Inhaberwechseln ist somit immer dann empfehlenswert, wenn der dadurch verursachte Aufwand für den jeweiligen Anwendungsbereich vertretbar ist.

12.1.4 Übersicht der Anreize zum Eingehen von Transaktionen

In den vorangehenden Abschnitten sind verschiedene weiterführende Konzepte vorgestellt worden. Sie setzen Versprechen ein, um Einheiten ein Anreiz zum Eingehen von Transaktionen zu geben, die für sie an sich unvorteilhaft sind. In diesem Abschnitt geben wir eine abschließende Übersicht der Anreize zum Eingehen von Transaktionen. Dabei werden die verschiedenen Arten von Anreizen (die *Anreizmuster*) untereinander in Beziehung gebracht.

Die Frage nach Anreizen zum Eingehen von Transaktionen ist nicht spezifisch für Informationssysteme. In der Geschichte der menschlichen Gesellschaft hat diese Frage ebenfalls eine große Rolle gespielt. Es lassen sich direkte Parallelen zwischen den Konzepten der vorangehenden Abschnitte und den Anreizmustern aufzeigen, die in der menschlichen Gesellschaft Verwendung fanden beziehungsweise immer noch finden [ON03]. Darauf weist unter Anderem die Namensgebung der Eigenwechsel und Inhaberwechsel hin, die aus dem realweltlichen Konzept der Wechsel entlehnt ist. Abstreitbare Versprechen im Sinne von Abschnitt 12.1.1 entsprechen den mündlichen Absprachen, die in kleinen abgeschlossenen Gesellschaften vorherrschen [And03a]. Die Art von Anreiz, die durch das Abgeben solcher Versprechen erzeugt wird, wird in [ON03] folglich *Gemeinschaftsmuster* (engl.: community pattern) genannt.

Abbildung 12.7 gibt eine Klassifikation der Anreizmuster, die wir identifiziert haben. Es gibt vier Arten von Anreizen, die eine Einheit zur Teilnahme an einer Transaktion bewegen können. Den *Tauschhandel* haben wir dem Hauptteil dieser Arbeit zugrunde gelegt. Sein Einsatz ist immer dann zu empfehlen, wenn der Transaktionspartner in der Lage ist, eine Aktion auszuführen, die von Interesse ist. So ist zum Beispiel in der ursprünglichen Beschreibung des Campus-Szenarios aus Abschnitt 1.2.1 der Einsatz von Versprechen überflüssig, da die Einheiten bei jeder Transaktion über jeweils passende Gegenleistungen verfügen.

Die Situation ändert sich allerdings, wenn zum Zeitpunkt der Transaktion keine Gegenleistung verfügbar ist. Dies ist zum Beispiel der Fall, wenn Anna auf den Mensaplan zugreifen will, bevor sie ihren Vorlesungsmitschrieb abgeschlossen hat und Claude anbieten kann. Dann bleibt als einzige Alternative, das zukünftige Erbringen einer Gegenleistung zu versprechen. Wenn Versprechen

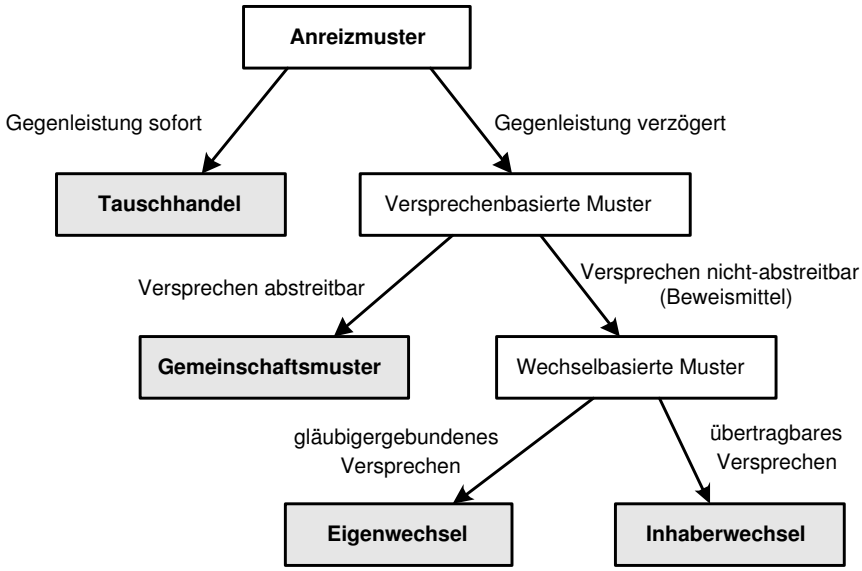

Abbildung 12.7: Einordnung der Anreize zum Eingehen von Transaktionen

abstreitbar sind, handelt es sich um das Gemeinschaftsmuster. Der Einsatz dieser Art von Anreiz ist allerdings nur dann zu empfehlen, wenn der Gläubiger großes Vertrauen in den Schuldner besitzt. Ansonsten sind aufgrund der relativ geringen Betrugskosten abstreitbare Versprechen mit zu großen Risiken verbunden.

Wir haben zwei Formen von nicht-abstreitbaren Versprechen vorgestellt: Eigenwechsel beinhalten Versprechen, die an den jeweiligen Gläubiger gebunden sind. Insbesondere ist das Anrecht auf die Einlösung eines Versprechens nicht übertragbar. Diese Einschränkung wird von Inhaberwechseln nicht geteilt. Daraus ergeben sich eine Reihe von Vorteilen gegenüber Eigenwechseln. Wenn die Umgebung des Informationssystems den Aufwand für die Verwaltung von Inhaberwechseln erlaubt, sind diese somit den Eigenwechseln vorzuziehen.

Die Überlegungen zeigen, dass die Wahl eines geeigneten Anreizmusters abhängig von den Eigenschaften der Anwendungsumgebung ist. Betrachtungen hierzu finden sich in [OKRK03].

Eine systematische Untersuchung der Anreizmuster in der menschlichen Gesellschaft wird in [ON03, AO03] gegeben. Darin werden einige weitere Alternativen zu den bisher eingeführten Anreizmustern aufgezeigt. Bei ihnen handelt es um das Anteil-, Bank- und Banknotenmuster. Diese Arten von Anreizen wurden in der Besprechung dieses Abschnitts nicht berücksichtigt, da sie auf den Einsatz von vertrauenswürdigen Dritten angewiesen sind. Diese sind aber in einem selbstorganisierenden Informationssystem nicht verfügbar. Es ist eine offene Fragestellung, ob diese Anreizmuster durch Anpassung ihrer Verfahren selbstorganisierend gemacht werden können. Einige Vorüberlegungen hierzu finden sich für das Anteil- und Bankmuster in [And03a].

12.2 Kooperationsanreize unter erschwerten Rahmenbedingungen

In diesem Abschnitt beschäftigen wir uns mit der Frage, ob der eigene Ansatz der verteilten Vertrauensbildung auch unter erschwerten Rahmenbedingungen einsetzbar ist. Im vorigen Abschnitt haben wir bereits einseitig vorteilhafte Transaktionen betrachtet. In diesem Abschnitt werden

drei weitere Erschwernisse untersucht:

- *Ungleiche Bedürfnisse und Fähigkeiten:* Hierbei sind zwei Fragestellungen von Interesse: Sollen Benutzer, die keinen Beitrag zum Informationssystem leisten können, in der Lage sein, dennoch am System teilzunehmen? Und wie ist die Teilnahme von solchen Benutzern zu gewährleisten, die zwar Beiträge leisten können aber keinen Bedarf an den im Informationssystem erhältlichen Informationen haben? Mit diesen Fragestellungen beschäftigt sich Abschnitt 12.2.1.
- *Mehrseitige Transaktionen:* Bislang sind wir davon ausgegangen, dass an jeder Transaktion stets zwei Einheiten beteiligt ist. Abschnitt 12.2.2 erörtert, warum dies in manchen Fällen nicht der Fall ist und wie mit mehrseitigen Transaktionen umzugehen ist.
- *Änderbare Identitäten:* Im Campus-Szenario besitzen die teilnehmenden Benutzer eine fest zugewiesene Identität. Allerdings sind auch andere Szenarien denkbar, in denen Identitäten änderbar sind. Mit den Problemen, die sich daraus ergeben, beschäftigt sich Abschnitt 12.2.3.

Die weiterführenden Konzepte, die zur Bewältigung dieser Erschwernisse anzuwenden sind und im Folgenden vorgestellt werden, bauen zum Teil auf den Überlegungen des vorangegangenen Abschnitts 12.1 auf. So bildet das Konzept zum Umgang mit einseitig vorteilhaften Transaktionen die Grundlage für die Konzepte, die ungleiche Bedürfnisse/Fähigkeiten und mehrseitige Transaktionen betreffen.

12.2.1 Ungleiche Bedürfnisse und Fähigkeiten

In diesem Abschnitt erörtern wir, welche Folgen sich ergeben, wenn die Benutzer und ihre Geräte in unterschiedlichem Maße vom Informationssystem abhängig sind und zu ihm beitragen können. Dazu diskutieren wir zunächst unterschiedliche Ausprägungen von Bedürfnissen und Fähigkeiten. Die Ergebnisse dieser Diskussion werden im Hinblick auf die Fairness zwischen den Einheiten erörtert. Es zeigt sich, dass auch die Fähigkeiten und Bedürfnisse zu berücksichtigen sind, die außerhalb des Informationssystems anzusiedeln sind. Als Substitut für fehlende Gegenleistungen im Informationssystem führen diese realweltlichen Aktionen zu vermehrter Kooperation im Gesamtsystem. Abschließend untersuchen wir, wie realweltliche Aktionen so gewichtet werden können, dass ihre Bedeutung auf angemessene Weise berücksichtigt wird.

Bedürfnisse und Fähigkeiten. Die Haltung einer Einheit im Bezug auf Kooperation wird von zwei Faktoren bestimmt. Auf der einen Seite steht das *Bedürfnis* einer Einheit, in Transaktionen einzugehen. Dieses ergibt sich daraus, dass andere Einheiten Aktionen ausführen können, die sie benötigt. Beispielsweise ist in der Szenariobeschreibung Bobs Einheit am Vorlesungsmitschrieb Annas interessiert und verspürt dadurch das Bedürfnis, in einer Transaktion mit Annas Einheit an diesen Mitschrieb zu gelangen. Auf der anderen Seite steht die *Fähigkeit* einer Einheit, Aktionen auszuführen, die von anderen Einheiten benötigt werden. Bobs Einheit bietet zum Beispiel den Konversionsdienst an, der Annas Einheit interessiert. Wie sich diese Faktoren auf das Kooperationsverhalten einer Einheit auswirken, zeigen wir im Folgenden an drei beispielhaften Belegungen dieser Faktoren.

Bislang sind wir von einem *Gleichgewicht zwischen Bedürfnis und Fähigkeit* einer Einheit ausgegangen. Dies bedeutet, dass eine Einheit in gleichem Maße der Kooperation bedarf, wie sie zu ihr fähig ist. In der Szenariobeschreibung aus Abschnitt 1.2.1 findet sich bei jeder Einheit

ein solches Gleichgewicht. Zum Beispiel ist Anna zwar am Mensaplan und am Konversionsdienst interessiert. Sie kann hierfür aber ihren Vorlesungsmitschrieb anbieten. Es spielt nur eine untergeordnete Rolle, ob die Menge der Einheiten, die am Vorlesungsmitschrieb interessiert sind, mit der Menge der Einheiten, die den Mensaplan und Konversionsdienst anbieten können, übereinstimmt. Wenn dem nicht der Fall ist, sorgt der Einsatz von den in Abschnitt 12.1.3 eingeführten Inhaberwechseln dafür, dass Annas Einheit dennoch zu den benötigten Informationen gelangt.

Die zweite beispielhafte Belegung der Faktoren ergibt sich aus *übergroßen Bedürfnissen*. Wenn die Bedürfnisse einer Einheit ihre Fähigkeiten übersteigen, so ergibt sich für sie ein Problem. Sie ist zwar an der Teilnahme in Transaktionen interessiert. Allerdings kann sie andere Einheiten aufgrund fehlender Fähigkeiten nicht zur Teilnahme bewegen. Daran ändern auch die Anreize aus Abschnitt 12.1 nichts. Der Mangel an Fähigkeiten führt nämlich dazu, dass keine Versprechen zukünftiger Gegenleistung gegeben werden können, die längerfristig als glaubwürdig erscheinen. Die Folge übergroßer Bedürfnisse ist also, dass eine Einheit nicht zu dem Maße am Informationssystem teilnimmt, wie sie dies gerne tun würde. Dies lässt sich durch eine Anpassung der Szenariobeschreibung illustrieren: Angenommen, Bob besitzt keinen Laptop sondern lediglich einen PDA und kann deswegen keinen Konversionsdienst mehr anbieten. Damit besitzt Bobs Einheit keine Fähigkeiten, die für andere Einheiten von Interesse sind. Als Folge davon gelingt es Bobs Einheit nicht, mit Annas Einheit in eine Transaktion einzugehen und darin den Vorlesungsmitschrieb zu erhalten.

Ein anderes Problem ergibt sich, wenn eine Einheit keine oder wenige Bedürfnisse besitzt. Es handelt sich somit um eine Einheit, die im Verhältnis zu ihren Bedürfnissen mit *übergroßen Fähigkeiten* ausgestattet ist. In diesem Fall besitzt sie keinen Anreiz dazu, in dem Maße am Informationssystem beizutragen, wie sie dazu fähig wäre. Wenn zum Beispiel Anna weder Mensaplan noch Konversionsdienst bräuchte, so würde die Teilnahme am Informationssystem für sie keinen Nutzen mit sich bringen. Folglich würde sie darauf verzichten, ihre Vorlesungsmitschriebe Anderen zur Verfügung zu stellen. Dem Informationssystem entgehen somit wichtige Quellen von Informationen.

Realweltliche Aktionen als Substitut. Die vorigen Überlegungen haben gezeigt, dass Einheiten mit übergroßen Bedürfnissen oder Fähigkeiten insofern ein Problem darstellen, als dadurch der Grad der Kooperation im Gesamtsystem eingeschränkt wird. Im Folgenden beurteilen wir, ob dieses Problem lösenswert ist und welche prinzipielle Lösung sich hierfür anbietet.

Auf den ersten Blick erscheint es so, als ob kein Handlungsbedarf besteht. Aufgrund der Autonomie der Benutzer kann eine Einheit mit übergroßen Fähigkeiten nicht dazu gezwungen werden, über einen längeren Zeitraum Transaktionen einzugehen, die für sie unvorteilhaft sind. Wenn dies im Systementwurf so vorgesehen würde, wäre normatives Verhalten nicht mehr hinreichend vorteilhaft und die Benutzer entschieden sich zur Manipulation. Andererseits ist es aus Gründen der Fairness problematisch, Einheiten mit übergroßen Bedürfnissen Transaktionen zuzugestehen, in denen sie nichts leisten. Sonst hätten die Benutzer keinen Anreiz, ihre Einheiten mit hinreichend vielen Fähigkeiten auszustatten. Zum Beispiel könnte Bob beabsichtigterweise zu einem PDA gewechselt haben, um nicht mehr in der Lage sein zu müssen, für Andere Konversionsdienste anbieten zu können. Das Fazit aus diesen Betrachtungen ist, dass aus der Sicht der Fairness eingeschränkte Kooperation sinnvoll und somit hinzunehmen ist.

Zu einem anderen Ergebnis kommen wir allerdings, wenn wir unsere Betrachtungen nicht auf das Informationssystem beschränken. Unter Umständen ist nämlich ein Benutzer in der Lage, in der Realwelt für andere Benutzer Aktionen auszuführen. Ein Beispiel hierfür lässt sich an einer Erweiterung der Szenariobeschreibung aufzeigen: Zwar ist Bobs Gerät nicht mehr zur Bereitstel-

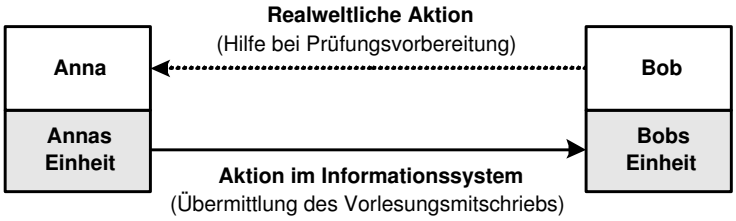

Abbildung 12.8: Beispiel einer Transaktion auf mehreren Ebenen

lung eines Konversionsdienstes fähig. Bob selbst könnte aber in der Realwelt Anna behilflich sein, zum Beispiel indem er sie bei der Vorbereitung auf eine Prüfung unterstützt. Somit ergibt sich eine Situation, die in Abbildung 12.8 dargestellt ist: Bobs Einheit besitzt übergroße Bedürfnisse, da sie Annas Einheit für die Bereitstellung des Vorlesungsmitschriebs keine Gegenleistung anbieten kann. Andererseits ist Bob in der Lage, eine solche Gegenleistung in der Realwelt zu erbringen. Auch aus Annas Sicht macht Kooperation Sinn. Zwar ist ihre Einheit an keiner Information im Rahmen des Informationssystems interessiert. Anna selbst könnte aber in der Realwelt durchaus Unterstützung gebrauchen. Kommt es zum Austausch von Vorlesungsmitschrieb und realweltlicher Hilfe, so nennen wir diese Transaktion eine *Transaktion auf mehreren Ebenen* (engl.: multi-layer transaction) [OK03b].

Das Ausführen realweltlicher Aktionen ist somit ein Substitut für fehlende Fähigkeiten im Informationssystem. Es ist wichtig zu betonen, dass die dabei zustande kommenden Transaktionen auf mehreren Ebenen aus Sicht des Gesamtsystems wünschenswert sind: Wäre die realweltliche Ebene von der Ebene des Informationssystems getrennt, so käme es zu keinem Austausch. Dies bezieht sich nicht nur darauf, dass Bobs Einheit nicht den benötigten Vorlesungsmitschrieb von Annas Einheit erhält. Darüber hinaus hätte Bob unter Umständen in der Realwelt keinen Anreiz, seine Zeit in die Prüfungsvorbereitung Annas zu investieren. Nur durch die ganzheitliche Betrachtung gelingt es, die zwischen Anna und Bob verborgenen Synergien auszunutzen.

Welche realweltlichen Aktionen sind denkbar? Bobs Hilfe bei Annas Prüfungsvorbereitung stellt nur einen Typus möglicher Aktionen dar. Darüber hinaus gibt es noch zwei prinzipielle Möglichkeiten, die beide für sich sinnvoll sind:

- *Monetäre Gegenleistung:* Für die Bereitstellung von Annas Vorlesungsmitschrieb könnte Bob sie ebenso gut mit realweltlichem Geld belohnen. Auch im universitären Umfeld ist diese Form der Entlohnung wünschenswert. Für Anna sichert das Anfertigen und Bereitstellen ihres Vorlesungsmitschriebs ein Nebeneinkommen[1].

- *Keine Gegenleistung:* Wie geht ein Benutzer vor, der seine Informationen bestimmten anderen Benutzern ohne Gegenleistung zur Verfügung stellen will? Der Systementwurf sieht dies nicht unmittelbar vor. Das Konzept der Transaktionen auf mehreren Ebenen ermöglicht jedoch einen solchen Altruismus. In diesem Fall könnten Benutzer untereinander vereinbaren, dass ihre Einheiten untereinander keine Gegenleistungen einfordern.

Es ist eine offene Fragestellung, wie das hier vorgebrachte Konzept von Transaktionen auf

[1] Aus den Erfahrungen des Autors besteht schon zurzeit diese Art der studentischen Aktivität. Am Anfang eines Semester kommt es vor der Vorlesungsräumen häufig dazu, dass Studenten höherer Semester ihre früheren Vorlesungsmitschriebe in nicht-elektronischer Form verkaufen. Die Möglichkeit monetärer Entlohnung ist somit auch im universitären Umfeld verankert.

mehreren Ebenen und realweltlichen Aktionen umgesetzt werden kann. Sie muss von zukünftigen Forschungsarbeiten angegangen werden.

Gebündelte Gegenleistungen für realweltliche Aktionen. Das Konzept realweltlicher Aktionen wirft ein Problem auf, das wir bisher noch nicht berücksichtigt haben: Da Bobs Einheit mit unzulänglichen Fähigkeiten ausgestattet ist, muss Bob für jede Transaktion seiner Einheit eine realweltliche Gegenleistung zur Verfügung stellen. Aus der Sicht Bobs ist somit ein aktives Eingreifen bei jeder Transaktion seiner Einheit notwendig. Dies bedeutet, dass die Teilnahme am Informationssystem nicht in dem Maße automatisiert erfolgt, wie es wünschenswert wäre.

Dieses Problem lässt sich unter Anwendung des Konzepts der Eigen- oder Inhaberwechsel aus Abschnitt 12.1 lösen. Eine realweltliche Aktion ist im Allgemeinen weitaus wertvoller als eine Aktion im Informationssystem. Im Gegenzug zur Ausführung einer realweltliche Aktion sollten somit im Informationssystem eine Reihe von Gegenleistungen erbracht werden. Hierzu bietet sich das Ausstellen oder Übertragen von Eigen- oder Inhaberwechseln an.

Wir verdeutlichen diesen Punkt durch ein Beispiel: Da Bob Zeit für Annas Prüfungsvorbereitung opfert, beauftragt sie im Gegenzug ihre Einheit, Bobs Einheit 10 Eigenwechsel auszustellen. Das Versprechen der Eigenwechsel besteht darin, jeweils einen zukünftigen Vorlesungsmitschrieb Bob zur Verfügung zu stellen. Mit Hilfe dieser Eigenwechsel ist Bobs Einheit in der Lage, an verschiedenen Zeitpunkten auf die benötigten Vorlesungsmitschriebe zuzugreifen, ohne dass eine Aktion seitens Bob erforderlich wird. Damit erfolgen aus der Sicht Bobs die Transaktionen seiner Einheit automatisiert.

Zusammenfassend erreicht das Konzept realweltlicher Aktionen das Zustandekommen von Kooperation, wann immer sie aus einer gesamtheitlichen Sicht sinnvoll erscheint. Obwohl hierfür explizite Handlungen seitens der Benutzer erforderlich sind, bleibt der Grad der Automatisierung unverändert hoch.

12.2.2 Mehrseitige Transaktionen

Bislang sind wir davon ausgegangen, dass an jeder Transaktion stets zwei Einheiten beteiligt sind. Dies ist zwar beim Austausch von Informationen der Fall. In diesem Abschnitt zeigen wir allerdings, dass im Hinblick auf Informationsdienste auch Transaktionen zwischen mehr als zwei Einheiten denkbar sind. Es gibt zwei Arten von mehrseitigen Transaktionen, die im Folgenden untersucht werden. Dabei handelt es sich zum einen um Transaktionen, in denen die Aktionen der Transaktionspartner zusammengesetzt sind. Zum anderen betrachten wir mehrseitige Kooperationsprotokolle, wie sie etwa in Vermittlungsoverlays wie Lanes (vergleiche Abschnitt 2.1) Verwendung finden. Diese Protokolle schreiben Abläufe der Kooperation vor, die nicht in einfache zweiseitige Transaktionen zerlegbar sind.

Zusammengesetzte Aktionen. Bei den bisherigen Überlegungen sind wir davon ausgegangen, dass die Einheiten ihre Aktionen ohne äußere Mithilfe ausführen können. In dieser Hinsicht sprechen wir von *einfachen Aktionen*. Beispielsweise ist in der Beschreibung des Campus-Szenarios aus Abschnitt 1.2.1 Annas Einheit in der Lage, ihren Vorlesungsmitschrieb anderen Einheiten übermitteln. Selbiges gilt für Bobs Einheit und seinen Konversionsdienst, den sie auf Anfrage für andere Einheiten ausführt.

Dass Aktionen nicht zwingendermaßen einfach sind, zeigt die folgende Erweiterung der Szenariobeschreibung: Nehmen wir einen Student Fabian an, der Annas Vorlesungsmitschrieb in einem abweichenden Dateiformat nutzen möchte. Dabei ist nur Bobs Gerät in der Lage, die dafür

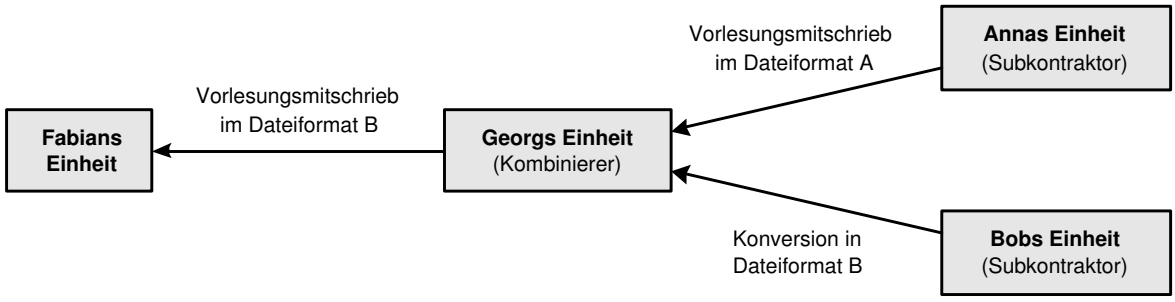

Abbildung 12.9: Beispiel für eine zusammengesetzte Aktion

notwendige Konversion auszuführen. Um den Vorlesungsmitschrieb in der gewünschten Form zu erhalten, benötigt Fabians Einheit somit die Mithilfe von zwei weiteren Einheiten. Dies lässt sich im Rahmen von zwei jeweils zweiseitigen Transaktionen erreichen: Zunächst erhält Fabians Einheit in einer ersten Transaktion von Annas Einheit den Vorlesungsmitschrieb. In einer zweiten Transaktion mit Bobs Einheit kommt es zu dessen Konversion. Im Zuge dieser Transaktionen belohnt Fabians Einheit Annas und Bobs Einheit mit je einem Eigenwechsel. Aus der Sicht von Fabians Einheit ist diese Vorgehensweise unter Umständen unbefriedigend. Wenn Bobs Einheit sie in der zweiten Transaktion um die Konversion betrügt, so war die erste Transaktion vergeblich. Fabians Einheit wird also nur dann in zwei separate Transaktionen einwilligen, wenn sie großes Vertrauen in Annas und vor allem Bobs Einheit besitzt. Ansonsten wird eine weitere Einheit benötigt, die für die Kombination der Aktionen von Annas und Bobs Einheit sorgt.

Abbildung 12.9 zeigt die sich ergebende Situation. Dabei ist der Student Georg der Prinzipal der kombinierenden Einheit. Fabians und Georgs Einheiten gehen in eine Transaktion, in der Georgs Einheit die Übermittlung von Annas Vorlesungsmitschrieb in dem von Fabian benötigten Dateiformat verspricht. Um dies tun zu können, geht Georgs Einheit zwei separate Transaktionen mit Annas und Bobs Einheiten ein. Bei der Aktion, die Georgs Einheit für Fabians Einheit ausführt, handelt es sich um eine *zusammengesetzte Aktion*. Aufgrund ihrer Rolle bei dieser Aktion nennen wir Annas und Bobs Einheit die *Subkontraktoren* und Georgs Einheit den *Kombinierer* der zusammengesetzten Aktion. Welchen Vorteil bietet diese Vorgehensweise im Vergleich zur Kombination durch Fabians Einheit selbst? Der Kombinierer der zusammengesetzten Aktion (in diesem Fall Georgs Einheit) kann nach dem Gesichtspunkt ausgewählt werden, dass er den Subkontraktoren vertraut. Damit ist also kein Vertrauen von Fabians Einheit in die Subkontraktoren vonnöten.

An sich lässt sich die Kooperation zwischen Annas, Bobs, Fabians und Georgs Einheit als drei jeweils zweiseitige Transaktionen beschreiben. Insofern sind keine Erweiterungen am Systemmodell vonnöten, das dem eigenen Ansatz der verteilten Vertrauensbildung zugrunde liegt. Andererseits stehen diese drei Transaktionen aber in einem semantischen Zusammenhang: Wenn Annas oder Bobs Einheit die Einheit Georgs betrügt, so kann Georgs Einheit ihre Aktion nicht wie verabredet ausführen. Aus der Sicht von Fabians Einheit erscheint dies wie ein Betrug durch Georgs Einheit. Aufgrund dieses semantischen Zusammenhangs sprechen wir von einer *mehrseitigen Transaktion* zwischen diesen vier Einheiten.

Wie ist der eigene Ansatz zu erweitern, um auf solche mehrseitigen Transaktionen eingehen zu können? Zunächst ist festzuhalten, dass eine solche Erweiterung nicht zwingend ist, wenn wir mehrseitige Transaktionen als eine Zusammensetzung mehrerer jeweils zweiseitiger Transaktionen betrachten. In diesem Fall nimmt der Kombinierer die Rolle eines Entrepreneurs ein, der für

etwaiges Betrugsverhalten der Subkontraktoren geradesteht. In gewissem Maße bürgt er somit für diese und wird für eine erfolgreiche Ausführung der zusammengesetzten Aktion eine besonders wertvolle Gegenleistung erwarten. Dabei ist eine offene Fragestellung, unter welchen Umständen eine Einheit bereit sein sollte, die Rolle des Kombinierers einzunehmen.

Eine Alternative zu dieser Sicht stellt eine echte Erweiterung des eigenen Ansatzes um mehrseitige Transaktionen dar. Dies erfordert eine Anpassung der Vorschriften der Glaubensrevision, da bei einem Fehler seitens des Kombinierers auch Betrug durch einen der Subkontraktoren in Erwägung gezogen werden muss. Wie diese Erweiterung durchzuführen ist, stellt ebenfalls eine offene Fragestellung dar, die in zukünftigen Forschungsarbeiten anzugehen ist.

Kooperationsprotokolle. Zusammengesetzte Aktionen stellen insofern ein überschaubares Problem dar, als sie sich als eine Zusammenführung mehrerer jeweils zweiseitiger Transaktionen verstehen lassen. Ein weitaus schwierigeres Problem ergibt sich jedoch dann, wenn die Kooperation zwischen mehreren Einheiten derart erfolgt, dass diese Interpretationsweise nicht möglich ist. Ein Beispiel für solche *Kooperationsprotokolle* sind die Protokolle, nach denen Overlays aufgebaut werden. Ein Beispiel für ein Overlay, das speziell für Ad-hoc Netze ausgerichtet ist, ist *Lanes* [KKRO03]. Im Folgenden wird die Problematik, die mit Kooperationsprotokollen verbunden ist, anhand dieses Overlays aufgezeigt. Um unnötige Wiederholungen zu vermeiden, werden in der nachfolgenden Betrachtung Information und Informationsdienst unter dem Begriff Information zusammengefasst.

Laut Abschnitt 2.1 dienen Overlays dazu, die Suche nach Informationen auf effiziente Weise zu unterstützen. Um dies zu tun, sieht das Lanes-Protokoll eine Overlay-Struktur gemäß Abbildung 12.10 vor: Die Einheiten sind als weiße Kästen dargestellt. Die Menge der am Lanes-Overlay teilnehmenden Einheiten ist in disjunkte Mengen aufgeteilt, die wir jeweils Lane nennen. In der Abbildung gibt es drei solcher Mengen. Sie sind grau hinterlegt. Innerhalb eine Lane geben die Einheiten in Form von *Angeboten* weiter, über welche Art von Informationen sie verfügen. Zu diesem Zweck sind die Einheiten einer Lane in einer linearen Struktur organisiert. Jede Einheit kennt ihren Nachfolger und Vorgänger innerhalb dieser Struktur[2]. Somit ist es möglich, dass Angebote entlang dieser linearen Struktur in einer Lane propagiert werden.

Die Kommunikation zwischen verschiedenen Lanes wird notwendig, wenn eine der Einheiten eine Information sucht und zu diesem Zweck eine entsprechende *Nachfrage* stellt. Ein *Vergleich* mit den Angeboten, von denen die Einheit im Zuge der Propagierung innerhalb der Lane erfahren hat, zeigt, ob eine der Einheiten der eigenen Lane über die benötigte Information verfügt. Wenn dem nicht so ist, fragt die Einheit bei einer benachbarten Lane an. Zu diesem Zweck kennt sie die Anycast-Adresse dieser Lane. Erhält eine Einheit eine solche Nachfrage, so untersucht sie durch lokalen Vergleich mit den bekannten Angeboten, ob in ihrer eigenen Lane eine entsprechende Information angeboten wird. Wenn sie fündig wird, gibt sie ein entsprechendes Ergebnis der nachfragenden Einheit zurück. Andernfalls leitet sie die Nachfrage an die jeweils benachbarte Lane weiter. Dadurch wird sichergestellt, dass durch die Nachfrage alle im Informationssystem verfügbaren Informationen gefunden werden.

Diese Beschreibung zeigt, dass im Lanes-Protokoll drei prinzipielle Arten von Aktionen vorgesehen sind:

1. *Teilnahme an Wartungsarbeiten:* Damit die beschriebene Lanes-Struktur zustande kommt

[2]Es gibt unterschiedliche Varianten des Lanes-Protokolls. Eine Variante, in der sogar Vor-Vorgänger bekannt sind, wird in [Pap03] vorgestellt.

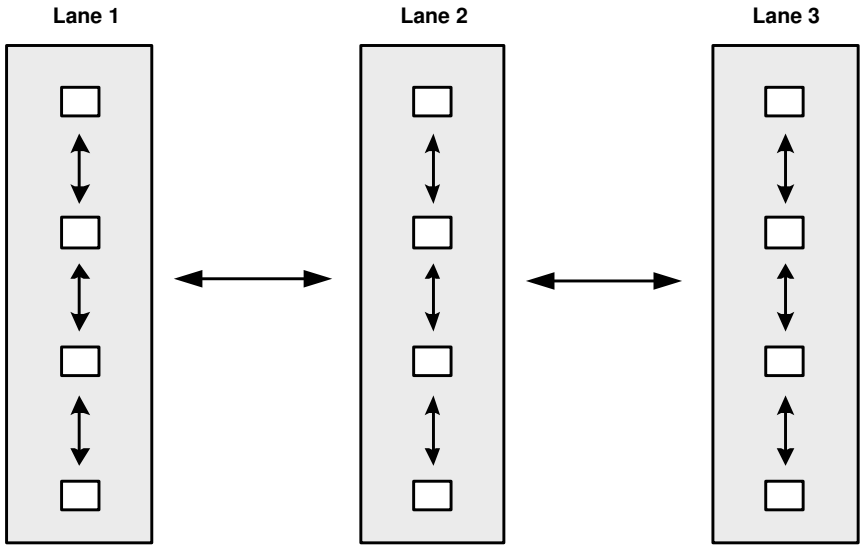

Abbildung 12.10: Schematische Darstellung der Struktur des Lanes-Overlays

und robust gegen Kommunikationsabbrüche ist, sieht das Protokoll bestimmte Wartungsarbeiten vor. Diese regulieren unter anderem die Größe einer Lane.

2. *Entgegennahme von Angeboten:* Werden Angebote innerhalb einer Lane propagiert, so sieht das Protokoll vor, dass jede Einheit der Lane diese Angebote bei sich lokal ablegt.

3. *Bearbeiten von Nachfragen:* Die Nachfrage wird mit den lokal abgelegten Angeboten verglichen. Im Erfolgsfall kommt es zur Benachrichtigung der nachfragenden Einheit und andernfalls zur Weiterleitung der Nachfrage.

Durch die Typisierung der im Lanes-Protokoll anfallenden Aktionen wird das prinzipielle Problem von Kooperationsprotokollen verdeutlicht: Gemäß dem Systemmodell aus Abschnitt 1.2.2 wird eine Aktion immer für genau eine Einheit (dem jeweiligen Transaktionspartner) ausgeführt. Dies ist im Lanes-Protokoll aber nur bei der dritten Art von Aktion der Fall. Bei ihr ist die nachfragende Einheit die Begünstigte der Aktionsausführung. Bei den ersten zwei Arten von Aktionen ist es hingegen unklar, wer von ihrer Ausführung profitiert. Damit wird deutlich, dass sich die Aktionen, die im Rahmen des Lanes-Protokolls ausgeführt werden, nicht auf das bisherige Transaktionsmodell zurückführen lassen.

Um das Lanes-Protokoll robust gegenüber Fehlverhalten zu machen, muss es um einen Ansatz der verteilten Vertrauensbildung erweitert werden. Die Beschäftigung mit dieser Aufgabe in [OKRP04, Pap03] bringt die folgenden Einsichten hervor:

- *Einsatzmöglichkeiten bisheriger Konzepte:* Mit Hilfe des Konzepts der Eigen- oder Inhaberwechsel aus Abschnitt 12.1 lässt sich ein Anreiz für die Bearbeitung von Nachfragen geben. So lässt sich zum Beispiel die erfolgreiche Bearbeitung entlohnen, indem die nachfragende Einheit einen entsprechenden Eigenwechsel zustellt. Bei anderen Arten von Aktionen ist dieser Ansatz allerdings nicht möglich, da es keine Einheit gibt, die sich als Begünstigter der Aktionsausführung identifizieren lässt.

- *Zielkonflikt zwischen Effizienz und Robustheit:* Der Entwurf von Kooperationsprotokollen wie Lanes orientiert sich an der Maximierung der Effizienz. Damit einher geht die implizi-

te Annahme, dass sich alle Einheiten entsprechend des Kooperationsprotokolls verhalten. Diese Annahme ist aber nicht haltbar, da die Benutzer manipulierte Versionen der Systemsoftware verwenden können. Dies zeigt, dass beim Entwurf von Kooperationsprotokollen sowohl Effizienz als auch Robustheit gegenüber Fehlverhalten zu berücksichtigen sind. Eine nachträgliche Erweiterung des Kooperationsprotokolls um robustheitssteigernde Maßnahmen gestaltet sich hingegen als äußerst schwierig. Beispielsweise stellt die Forderung, dass Nachfragen an benachbarte Lanes weitergeleitet werden sollen, ein gewisses Problem dar. Die nachfragende Einheit kennt nämlich aufgrund des Anycast-Mechanismus die Einheit nicht, die für diese Weiterleitung zuständig ist. Unterbleibt die Weiterleitung, so ist unklar, welche Einheit hierfür zur Verantwortung zu ziehen ist.

- *Abgrenzung zwischen Norm und Freiheit:* Gemäß der Ausrichtung des normativen Systementwurfs aus Abschnitt 5.4.2 bezieht sich soziale Kontrolle nur auf Norm-bezogenes Verhalten. In anderen Verhaltensbereichen sind die Einheiten hingegen frei in der Wahl ihres Verhaltens. Im Systementwurf aus Teil II dieser Arbeit schlägt sich dies zum Beispiel dadurch nieder, dass Einheiten nicht bestraft werden, wenn sie das Eingehen von Transaktionen verweigern. Die Abgrenzung zwischen Norm und Freiheit ist um ein vielfaches schwieriger, wenn wir statt zweiseitiger Transaktionen Kooperationsprotokolle wie Lanes betrachten. So ist etwa die Teilnahme an Wartungsarbeiten unbedingt erforderlich, damit die Effektivität des Overlays gewährleistet bleibt. Insofern müsste die Teilnahme an Wartungsarbeiten in einer entsprechenden Norm festgelegt sein. Andererseits gibt es keine Einheit, die die Einhaltung dieser Norm überprüfen kann, geschweige denn andere Einheiten darüber glaubwürdig informieren könnte. Hier zeigt sich wiederum, dass die nachträgliche Erweiterung von Kooperationsprotokollen um Robustheit äußerst schwierig ist.

Diese Betrachtungen zeigen, dass die Anwendung der verteilten Vertrauensbildung auf Kooperationsprotokolle wie Lanes mit erheblichen Schwierigkeiten verbunden ist. Wie diese Schwierigkeiten zu überwinden sind, stellt eine offene Fragestellung dar, die in zukünftigen Forschungsarbeiten zu beantworten ist. Unter Umständen bieten sich hierfür die Theorien der Externalität und der öffentlichen Güter als Ausgangspunkt an [MR04]. Diese aus den Wirtschaftswissenschaften kommenden Theorien befassen sich damit, wie Aktionen zu motivieren sind, deren Nutzen nicht eindeutig einer Einheit zuordenbar ist.

Eine wichtige Schlussfolgerung der Überlegungen dieses Abschnitts bezieht sich auf den Entwurf von Kooperationsprotokollen: Dieser muss nicht nur im Hinblick auf Effizienz sondern auch auf Robustheit gegenüber Fehlverhalten erfolgen, da nachträgliche Erweiterungen äußerst schwierig oder unmöglich sind. Bereits beim Entwurfszeitpunkt ist somit die Fragestellung zu berücksichtigen, welches Verhalten für die Effektivität des Protokolls unabdingbar ist und welches nicht. Für die Effektivität unabdingbares Verhalten muss nicht nur in entsprechenden Normen festgelegt werden sondern auch durch angemessene Maßnahmen kontrolliert und in der Vertrauensbildung berücksichtigt werden. Freiwilliges Verhalten erfordert hingegen die Schaffung entsprechender Anreize etwa mit Hilfe des Konzepts der Eigen- oder Inhaberwechsel.

12.2.3 Änderbare Identitäten

Im Campus-Szenario besitzen die teilnehmenden Benutzer eine fest zugewiesene Identität. Allerdings sind auch andere Szenarien denkbar, in denen Identitäten änderbar sind. In diesem Abschnitt beschäftigen wir uns mit den Problemen, die sich daraus ergeben. Hierzu gehen wir auf die Eigenschaften von änderbaren Identitäten ein. Außerdem werden die Auswirkungen auf die ver-

teilte Vertrauensbildung untersucht. Abschließend ziehen wir ein Fazit, wie änderbare Identitäten aus der Sicht des Systementwurfs zu bewerten sind.

Eigenschaften. Bisher sind wir davon ausgegangen, dass jede Einheit eine eindeutige Identität besitzt, die sie im Laufe der Zeit nicht abändern kann. Als Grundlage hierfür dient der kryptographische Ansatz, der in Abschnitt 2.1 beschrieben wurde. Er setzt voraus, dass jeder Benutzer initial ein Zertifikat seiner Identität zugewiesen bekommt. Im Campus-Szenario bietet sich hierfür eine zentrale Einrichtung der Universität (etwa das Rechenzentrum, das Studentenwerk oder die Universitätsbibliothek) an. Diese zentrale Einrichtung ist in der Lage, sich der Identität des Studenten zu vergewissern, indem sie das Vorzeigen seines Personal- oder Studentenausweises für die Zertifizierung seiner Identität voraussetzt. Speziell im universitären Umfeld ist dies unter Umständen nicht nötig, da die Universität aus anderen verwaltungstechnischen Gründen die Identität der Studenten überprüfen muss. Somit ist im Campus-Szenario die Annahme sinnvoll, dass Identitäten unabänderbar sind.

In anderen Szenarien ist dies jedoch nicht zwingendermaßen der Fall. Im Extremfall gibt es keine zentrale Instanz zum initialen Ausstellen von Zertifikaten der Benutzeridentität. Dann müssen die Benutzer ihre Identität selbst zertifizieren. Damit sind die Kosten für das Ändern der eigenen Identität sehr gering, da sie nur das Ausführen einer kryptographischen Operation erfordert [DA04]. Realistischer ist in den allermeisten Szenarien jedoch, dass es zwar eine zentrale Instanz zum Ausstellen von Zertifikaten gibt, diese aber nicht zusichern kann, dass jeder Benutzer maximal eine Identität zertifiziert bekommt. Als Beispiel hierfür lässt sich die Zertifizierung einer Identität aufgrund des Besitzes einer Kreditkarte anführen [Del03]. In diesem Fall kann ein Benutzer durch den Erwerb mehrerer Kreditkarten zwar zu mehreren Identitäten gelangen. Allerdings wird dies in zweierlei Hinsicht erschwert. Zum einen ist der Erwerb einer Kreditkarte mit nicht unerheblichen Kosten verbunden. Zum anderen ist es für einen Benutzer mit einem beträchtlichen Aufwand verbunden, ein und dieselbe Kreditkartenfirma vom Ausstellen mehrerer Kreditkarten zu überzeugen [Del03]. Somit ist die Zahl der Identitäten, die ein Benutzer zertifiziert bekommen kann, beschränkt.

Die Kosten der Identitätsänderung rühren nicht nur vom Überwinden technischer Hindernisse. Hinzu kommen rechtliche Hindernisse. Abschnitt 5.2.2.3 zeigt, dass die Zertifizierung einer Identität mit der Einwilligung in eine Lizenzvereinbarung verbunden werden kann. Wenn diese das Führen verschiedener Identitäten ausschließt, stellt das Ändern der eigenen Identität ein Verstoß gegen das Vertragsrecht dar. In Analogie zu den Hindernissen zur Manipulation aus Abschnitt 5.2 sind somit Identitätsänderungen in jedem Fall mit gewissen Kosten verbunden.

Die Änderbarkeit von Identitäten führt zu einer Erweiterung des Systemmodells. Die Einheit ein und desselben Benutzers kann sich nunmehr unter verschiedenen Identitäten ausgeben. Dabei ist es für andere Einheiten nicht wahrnehmbar, dass sich hinter den ausgegebenen Identitäten dieselbe Einheit verbirgt. Vielmehr erscheint es so, als ob es sich um unterschiedliche Einheiten handelt. Aus der Sicht des Gesamtsystems nimmt der Benutzer somit mit mehreren Einheiten teil. Diese nennen wir die *virtuellen Einheiten* des Benutzers, da er in Wirklichkeit lediglich über eine Einheit verfügt.

Auswirkungen auf die verteilte Vertrauensbildung. Im Folgenden gehen wir der Fragestellung nach, wie sich die Änderbarkeit der Identitäten auf die Effektivität des eigenen Ansatzes der verteilten Vertrauensbildung auswirkt. Zu diesem Zweck müssen wir untersuchen, wie ein Benutzer den Besitz mehrerer Identitäten zu seinem Vorteil ausnutzen kann.

Ein offensichtliche Möglichkeit besteht darin, dass sich die virtuellen Einheiten eines Benutzers gegenseitig hochloben. Dabei stellt sich allerdings die Frage, wie dies geschehen kann. Laut Abschnitt 2.4 sind in den konventionellen Ansätzen der verteilten Vertrauensbildung Empfehlungen als billiges Gerede anzusehen. Da sich dort zudem Einheiten auch positiv empfehlen können, leiden diese Ansätze darunter, gegenseitigem Hochleben nichts entgegenstellen zu können. Anders verhält es sich jedoch im eigenen Ansatz. Das Empfehlungssystem aus Abschnitt 7.5 sieht die Selbstempfehlung als einzige Art der positiven Empfehlung vor. Gemäß Abschnitt 7.4 und 8.3 kann sie nur dann zur Aufwertung des Selbstempfehlers führen, wenn dieser Bürgen besitzt und in der Selbstempfehlung diese angibt. Gegenseitiges Hochloben zwischen den virtuellen Einheiten desselben Benutzers besteht daher zwingendermaßen im Eingehen von Bürgschaftsbeziehungen untereinander. Aufgrund des Investitionsaspekts von Bürgschaften ist dies aber nur dann zum Vorteil des Benutzers, wenn sich seine virtuellen Einheiten kooperativ verhalten. Ansonsten führen die Revisionsvorschriften aus Abschnitt 8.3.2 dazu, dass seine virtuellen Einheiten kollektiv abgewertet werden. Somit ist gegenseitiges Hochloben ineffektiv darin, die Folgekosten für Betrugsverhalten zu verringern (vergleiche [OFN04]). Zusammenfassend ist der eigene Ansatz also robust gegenüber der Möglichkeit gegenseitigen Hochlobens.

Welchen echten Nutzen kann das Führen mehrerer Identitäten also mit sich bringen? Nehmen wir einen Benutzer an, dessen Einheit durch das Ausstellen inkonsistenter Beweismittel von anderen Einheiten als strategisch erkannt worden ist. Der Typglaube der anderen Einheiten über seine Einheit ist allerdings an die Identität gebunden, mit der sich seine Einheit ausgegeben hat. Wechselt die Einheit ihre Identität, so erscheint sie den anderen Einheiten als ein Neuankömmling. Gemäß Abschnitt 6.3.3 ergibt sich der Typglaube über sie somit direkt aus dem Systemglauben. Als Folge ihrer Identitätsänderung wird die Einheit also nicht mehr als sicher strategisch angesehen und kommt unter Umständen zu weiteren Transaktionen. Dieses Beispiel zeigt, dass die Möglichkeit der Identitätsänderung zur *Verringerung der Betrugsfolgekosten* führt. Damit einher geht eine Entschärfung der sozialen Kontrolle. Dies ist ein inhärentes Problem der verteilten Vertrauensbildung, das sich unabhängig vom gewählten Ansatz ergibt.

Die Auswirkungen dieses Sachverhalts auf das Gesamtsystem sind von der Höhe der Kosten für die Identitätsänderung abhängig. Wenn diese Kosten vernachlässigbar sind, wird eine Einheit immer dann ihre Identität wechseln, wenn der Typglaube anderer Einheiten über sie niedriger als ihr jeweiliger Systemglaube ist. Langfristig ergibt sich daraus eine ständige Verminderung des Systemglaubens. Dadurch wird die *Offenheit* des Informationssystems gefährdet. Dies wird deutlich, wenn wir die Sicht eines Benutzers annehmen, der sich gerade zur Teilnahme am Informationssystem entschieden hat. Seine Einheit kann den anderen Einheiten nicht glaubhaft vermitteln, dass sie keine strategische Einheit ist, die soeben ihre Identität geändert hat. Somit wird seine Einheit von Anderen als eher strategisch eingeschätzt und faktisch von Transaktionen ausgeschlossen. Das Informationssystem degeneriert also zu einem geschlossenen System, in dem Neuankömmlinge keine oder kaum eine Chance haben, sich zu etablieren.

Dieser Effekt tritt in dieser Schärfe allerdings nur dann zu Tage, wenn die Kosten der Identitätsänderung sehr gering sind. Wie bereits dargelegt ist dies aber in den allermeisten Szenarien nicht der Fall. Die nicht zu vernachlässigenden Kosten der Identitätsänderung führen dort dazu, dass Einheiten sehr viel seltener ihre Identität ändern wollen. Es ist eine offene Fragestellung, wie hoch die Kosten der Identitätsänderung mindestens sein müssen, damit die Offenheit des Informationssystems gewährleistet bleibt.

Fazit. Im Bezug auf änderbare Identitäten beschränkt sich das Wirkungsfeld der verteilten Vertrauensbildung darauf, gegenseitiges Hochloben unvorteilhaft zu machen. Wir haben gezeigt,

dass der eigene Ansatz im Gegensatz zu den anderen existierenden Ansätzen dafür sorgt.

Es ist ein prinzipielles Problem der verteilten Vertrauensbildung, dass die Änderbarkeit von Identitäten die Offenheit des Systems gefährdet. Diese Gefahr ist umso größer je geringer die Kosten sind, die mit einer Identitätsänderung verbunden sind. Es ist daher die Aufgabe des Systementwerfers, durch entsprechende Maßnahmen diese Kosten so hoch wie möglich zu halten.

12.3 Erweiterungsmöglichkeiten des Entwurfes und der Evaluation

In diesem Abschnitt befassen wir uns mit möglichen Erweiterungsrichtungen des Ansatzes dieser Arbeit. Hierfür werden Konzepte vorgestellt, die Aspekte des Entwurfs und der Evaluation sinnvoll ergänzen. Konkret werden zwei Erweiterungsmöglichkeiten aufgezeigt: Abschnitt 12.3.1 bespricht Alternativen zum Sechs-Wege Transaktionsprotokoll. Es zeigt sich, dass eine Änderung des Transaktionsprotokolls Folgen für die Glaubensbildung und das Empfehlungssystem mit sich bringt. Abschnitt 12.3.2 befasst sich mit der Frage, ob sich unabhängig vom Systementwurf alternative Normen als Folge der Manipulation etablieren können. Dabei wird besprochen, wie eine Antwort auf diese Frage im Rahmen der Evaluation zu finden ist.

12.3.1 Alternative Transaktionsprotokolle

Im Zuge des Systementwurfs haben wir in Abschnitt 7.2.2 die Verwendung eines Sechs-Wege Transaktionsprotokolls vorgeschlagen. In diesem Abschnitt beschäftigen wir uns damit, ob es Alternativen zu diesem Protokoll gibt. Dazu gehen wir zunächst auf die Anforderungen ein, die ein Transaktionsprotokoll erfüllen muss. Ausgehend von diesen Anforderungen identifizieren wir zwei sinnvolle Alternativen zum Sechs-Wege Transaktionsprotokoll. Bei jedem dieser beiden Protokolle besprechen wir, welche Folgen ihr Einsatz für den eigenen Ansatz der verteilten Vertrauensbildung hat.

Anforderungen. Um Alternativen zum Sechs-Wege Protokoll identifizieren zu können, müssen wir uns zunächst damit auseinander setzen, welche Eigenschaften wir vom Transaktionsprotokoll fordern. Im Folgenden werden zwei solcher Anforderungen erörtert.

Das Sechs-Wege Protokoll erweitert das Zwei-Wege Protokoll aus Abschnitt 6.2 um den Austausch von Verträgen und Quittungen. Der Einsatz dieser transaktionalen Beweismittel ist für die Funktionsweise des Empfehlungssystems unabdingbar. Daher müssen wir fordern, dass das Transaktionsprotokoll für den Austausch der Verträge und Quittungen sorgt. Insbesondere bedeutet dies, dass weder auf Verträge noch auf Quittungen verzichtet werden darf. Werden zum Beispiel nur Verträge aber keine Quittungen ausgetauscht, sind die beiden Transaktionspartner auch nach einer erfolgreichen Transaktion durch negative Empfehlungen angreifbar. Umgekehrt darf auf Verträge nicht verzichtet werden, da ohne sie keine negativen Empfehlungen möglich sind.

Die zweite Anforderung betrifft den Ablauf des Transaktionsprotokolls. In Abbildung 7.5 wurden die Phasen einer Transaktion dargestellt. Demgemäß muss der Austausch von Verträgen vor und der Austausch von Quittungen nach den Aktionsausführungen erfolgen. Alternative Protokolle dürfen diesen phasenorientierten Ablauf nicht abändern.

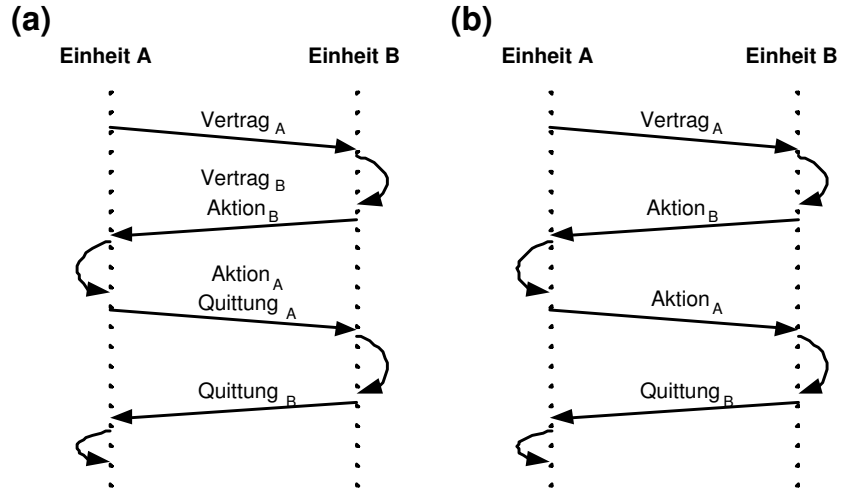

Abbildung 12.11: Alternative Transaktionsprotokolle: **(a)** das Vier-Wege Protokoll und **(b)** das asymmetrische Protokoll

Alternativen zum Sechs-Wege Protokoll. Auf den ersten Blick erscheint es so, als ob das Sechs-Wege Protokoll die einzige Möglichkeit darstellt die beiden Anforderungen zu erfüllen. Im Folgenden zeigen wir, dass dem nicht so ist. Hierfür werden zwei alternative Transaktionsprotokolle aufgezeigt [ON04].

Die erste Alternative ist das *Vier-Wege Transaktionsprotokoll*, das in Abbildung 12.11(a) dargestellt ist. Es unterscheidet sich vom Sechs-Wege Protokoll in einem einzigen Punkt: Einheit A führt ihre Aktion erst dann aus, wenn Einheit B ihre Aktion ausgeführt hat. Durch diese Umstellung im Protokollablauf sind beide Anforderungen weiterhin erfüllt. Durch die Zusammenfassung von Protokollschritten ergibt sich jedoch eine Ersparnis von zwei Nachrichten: Einheit B sendet das Ergebnis ihrer Aktionsausführung zusammen mit ihrem Vertrag. Analog dazu fügt Einheit A dem Ergebnis ihrer Aktionsausführung ihre Quittung hinzu.

Wie ist das Vier-Wege Protokoll zu bewerten? Im Vergleich zum Sechs-Wege Protokoll wird das Verschicken von insgesamt zwei Nachrichten eingespart. Dies führt nicht nur zu einem geringen Aufwand sondern auch zu einer Zeitersparnis. Allerdings bringt das Vier-Wege Protokoll mit sich, dass die Rollen zwischen den beiden Transaktionspartnern nicht mehr klar aufgeteilt sind: Im Bezug auf die Ausstellung transaktionaler Beweismittel trägt Einheit A die Risiko-Position, während sie bezüglich der Aktionsausführung die sichere Position inne hält. Diese Vermischung der Positionen ist allerdings insofern unkritisch, als sie keine Erweiterung des eigenen Ansatzes erfordert. Laut Abschnitt 6.4 hängt das Treffen von Entscheidungen zum Eingang in Transaktionen nämlich lediglich von der Position im Bezug auf die Aktionsausführung ab. Somit lassen sich die bisherigen Vorschriften auf das Vier-Wege Protokoll übertragen.

Die zweite Alternative ergibt sich aus der Überlegung, wie sinnvoll die Schritte des Vier-Wege Protokolls sind. Genauer gesagt ist es fragwürdig, warum Einheit B ihren Vertrag_B zusammen mit dem Ergebnis ihrer Aktionsausführung übermittelt. Der Vertrag dient dazu, dass Einheit A sie negativ empfehlen kann, wenn die Ausführung ihrer Aktion_B unterbleibt. Wenn Einheit B Betrugsverhalten beabsichtigt, würde sie somit in ihrem eigenen Interesse nicht nur die Ausführung der Aktion sondern auch das Ausstellen des Vertrages unterlassen. Wir erhalten also, dass Einheit A im Betrugsfall Einheit B gar nicht negativ empfehlen kann und der Vertrag_B somit wertlos ist. Selbiges gilt für die Quittung_A, die Einheit A zusammen mit dem Ergebnis ihrer

Tabelle 12.1: Parallele zwischen dem asymmetrischen Transaktionsprotokoll und dem Konzept der Eigenwechsel

Transaktionsprotokoll	Vertrag_A	Quittung_B	Einheit A	Einheit B
Eigenwechsel	Eigenwechsel	Entwertung	Schuldner	Gläubiger

Aktionsausführung übermittelt. Wenn wir nicht länger die Ausstellung der im Vier-Wege Protokoll überflüssigen Beweismittel fordern, erhalten wir das *asymmetrische Transaktionsprotokoll*, das in Abbildung 12.11(b) dargestellt ist.

Welche Eigenschaften besitzt dieses Protokoll? Es unterscheidet sich vom Vier-Wege Protokoll nur durch das Weglassen von Vertrag_B und Quittung_A, wodurch der Aufwand der Protokollausführung weiter gesenkt wird. Auf der anderen Seite wird aber auch deutlich, dass dieses Weglassen die erste der beiden Anforderungen an das Transaktionsprotokoll zum Teil verletzt. Dies hat zur Folge, dass Einheit B nicht mehr negativ empfohlen werden kann. Einheit A ist somit nicht mehr gegenüber Betrugsverhalten der Einheit B abgesichert. Die Benennung dieses Transaktionsprotokolls bezieht sich eben auf diese asymmetrische Verteilung der Angreifbarkeit zwischen den beiden Transaktionspartnern. Dies hat eine direkte Auswirkung auf das Empfehlungssystem. Da nur einer der beiden Transaktionspartner negativ empfehlen kann, verfällt der komparative Vorteil des Erstempfehlens. Gemäß Abschnitt 7.6.2 haben die Einheiten somit keinen Anreiz für das Ausstellen negativer Empfehlungen. Das asymmetrische Transaktionsprotokoll erfordert also Anpassungen im Empfehlungssystem und unter Umständen in der Glaubensbildung.

Tabelle 12.1 zeigt die Parallelen auf, die zwischen dem asymmetrischen Transaktionsprotokoll und dem Konzept der Eigenwechsel aus Abschnitt 12.1.2 bestehen. Die ersten beiden Schritte des Transaktionsprotokolls entsprechen der initialen Transaktion, in der Einheit A als der Schuldner durch Ausstellung eines entsprechenden Versprechens (der Vertrag beziehungsweise der Eigenwechsel) Einheit B zur Ausführung ihrer Aktion bringt. Der dritte und vierte Schritt des Transaktionsprotokolls entsprechen der Einlösungstransaktion, in der Einheit B als der Gläubiger Einheit A das Einlösen ihres Versprechens durch die Entwertung (beziehungsweise durch die Quittung) bescheinigt. Die Existenz dieser Parallele erklärt, warum sowohl beim asymmetrischen Transaktionsprotokoll als auch beim Konzept der Eigenwechsel eine Erweiterung des Empfehlungssystems erforderlich wird. Wenn in zukünftigen Forschungsarbeiten zum Konzept der Eigenwechsel eine solche Erweiterung vorgenommen wird, lässt sich der dabei angewendete Lösungsansatz somit auch auf das asymmetrische Transaktionsprotokoll übertragen.

12.3.2 Umgang mit abweichenden Normen

Der eigene Ansatz der verteilten Vertrauensbildung ist das Ergebnis des normativen Systementwurfs aus Teil II dieser Arbeit. Als solcher besteht er aus einer Reihe von Normen und Vorschriften, die Eingang in die originale Systemsoftware finden. Manipulierte Versionen dieser Systemsoftware müssen sich nicht notwendigerweise an diese Normen und Vorschriften halten. Dies haben wir im Abschnitt 10.1 beobachten können. Die dort aufgefundenen Gegenstrategien sehen unter ganz bestimmten Bedingungen Norm-verletzendes Verhalten (also Betrug im Zuge einer Transaktion) vor. Beim Finden dieser Gegenstrategien sind wir davon ausgegangen, dass sich zwischen strategischen Einheiten keine Normen entwickeln, die vom Systementwurf abweichen. Laut Abschnitt 5.3.3 könnte eine solche abweichende Norm etwa darin bestehen, dass sich die Einheiten, die dieselbe manipulierte Version der Systemsoftware benutzen, gegenseitig unterstützen.

In diesem Abschnitt erörtern wir, ob die Herausbildung solcher abweichender Normen zu erwarten ist und wie eine Antwort hierauf im Zuge der Evaluation gefunden werden kann. Zu diesem Zweck nehmen wir zunächst die Sicht eines Benutzers an, der eine manipulierte Version der Systemsoftware erstellt, und untersuchen den Entwurfsraum, der ihm bei der Definition abweichender Normen zur Verfügung steht. Basierend auf den dabei gewonnenen Erkenntnissen analysieren wir, ob zu erwarten ist, dass sich diese Normen im Gesamtsystem durchsetzen. Abschließend gehen wir darauf ein, wie die Evaluationsmethodik zu erweitern ist, um die Möglichkeit der Norm-setzenden Manipulation zu berücksichtigen.

Entwurfsraum für Normen-setzende Manipulation. Laut Abschnitt 5.3.3 setzen sich nur solche manipulierte Versionen der Systemsoftware durch, die individuell rationales Verhalten vorsehen. Das bedeutet, dass die Einheiten, die diese Versionen benutzen, ihr Verhalten allein am Vorteil ihrer jeweiligen Benutzer ausrichten. Die individuelle Rationalität von strategischen Einheiten hat zur Folge, dass sich strategische Einheiten untereinander betrügen können, auch wenn ihre jeweiligen Prinzipale dieselbe manipulierte Version der Systemsoftware verwenden. Auf den ersten Blick liegt es daher für den Ersteller einer manipulierten Version nahe, dafür zu sorgen, dass die Einheiten der übernahmewilligen Benutzer sich gegenseitig unterstützen und somit ein Komplott bilden. Für die Festlegung des Verhaltens der Komplottmitglieder muss der Ersteller auf zwei Punkte eingehen:

- *Verhalten gegenüber Mitgliedern:* Hierbei ist zu bestimmen, inwiefern sich die Mitglieder eines Komplotts untereinander unterstützen. Eine offensichtliche Möglichkeit besteht darin, dass sich Komplottmitglieder untereinander nicht betrügen.

- *Verhalten gegenüber Nichtmitgliedern:* Gegenüber Einheiten, die nicht Mitglied im Komplott sind, sind Komplottmitglieder eher zum Betrug bereit. Sie können sich unter gewissen Umständen aber auch kooperativ verhalten.

Die Festlegung dieser beiden Punkte stellt im Sinne von Abschnitt 5.4.2 die Definition von Vorschriften dar. Um Normen würde es sich hierbei nur handeln, wenn die Komplottmitglieder ihr Verhalten bezüglich dieser zwei Punkte untereinander kontrollieren.

Dass gegenseitige Kontrolle zum Funktionieren eines Komplotts erforderlich ist, zeigt die Darstellung aus Abschnitt 5.3.3: Sei V die manipulierte Version der Systemsoftware, in der die Herausbildung des Komplotts vorgesehen ist. Damit sich die Einheiten, deren Prinzipale diese Version V verwenden, untereinander als Komplottmitglieder zu erkennen geben können, bedarf es eines entsprechenden Erkennungsmechanismus. Nur so ist es einem Komplottmitglied möglich, sein Verhalten anderen Einheiten gegenüber an deren Zugehörigkeit zum Komplott auszurichten. Allerdings kann der Erkennungsmechanismus auch in einer weiteren manipulierte Version V' integriert werden. Somit können sich die Einheiten der Version V' als Komplottmitglieder ausgeben (so genannte *unechte Komplottmitglieder*). Dies ist insofern problematisch, als diese Einheiten von den Vorzügen der Komplottmitgliedschaft profitieren, ohne notwendigerweise den Verpflichtungen gegenüber den anderen Komplottmitgliedern nachzukommen. Somit ist die Verwendung der Version V' vorteilhafter als die der Version V und das Komplott bricht zusammen. Als Folge davon reicht ein Erkennungsmechanismus nicht aus. Zusätzlich müssen die Komplottmitglieder ihr Verhalten bezüglich dieser zwei obigen Punkte untereinander kontrollieren. Hierzu sind drei weitere Punkte im Entwurf der Version V zu berücksichtigen:

- *Definition von Normen:* Es ist festzulegen, welches Verhalten von Komplottmitgliedern erwartet und kontrolliert wird. Zum Beispiel könnte eine Norm sein, dass sich Komplottmit-

glieder nicht gegenseitig betrügen wollen. Die hierbei festgelegten Normen unterscheiden sich in jedem Fall von denen des eigenen Ansatzes der verteilten Vertrauensbildung, da sie sich nur auf Komplottmitglieder und nicht (wie im eigenen Ansatz) auf alle Einheiten beziehen.

- *Soziale Kontrolle der Komplottmitglieder:* Die Komplottmitglieder müssen sich im Bezug auf Norm-bezogenes Verhalten gegenseitig kontrollieren. Nur so können Einheiten der Version V' erkannt werden, die sich zwar als Komplottmitglieder ausgeben, aber andere Komplottmitglieder absichtlich betrügen. In Analogie zum eigenen Ansatz umfasst die soziale Kontrolle auch Verfahren der Glaubensbildung. Jedes Komplottmitglied bildet sich seinen Glauben darüber, mit welcher Wahrscheinlichkeit es sich bei einer Einheit, die sich als Komplottmitglied ausgibt, um ein echtes Mitglied handelt.

- *Verhalten gegenüber unechten Komplottmitgliedern:* Erscheint ein angebliches Komplottmitglied als unechtes Mitglied, so stellt sich die Frage, wie mit ihm umzugehen ist. Auch in diesem Punkt besteht eine Parallele zum eigenen Ansatz: Auch ein echtes Komplottmitglied könnte durch unbeabsichtigtes Betrugsverhalten fälschlicherweise als unechtes Mitglied erkannt werden. Dies ist bei der Beantwortung dieser Frage zu berücksichtigen.

Die Besprechung zeigt, dass diese Punkte vom Ersteller der Version V analog zum Ansatz dieser Arbeit bearbeitet werden können. Als Ergebnis entsteht im Informationssystem ein Komplott, dessen Mitglieder sich nach Normen richten, die spezifisch für das Komplott sind.

Selbst-Durchsetzung. Die bisherigen Betrachtungen bezogen sich auf den Gestaltungsspielraum von Komplotten und ihren vom Systementwurf abweichenden Normen. Im Folgenden beschäftigen wir uns mit der Frage, ob die Herausbildung solcher Komplotte und damit die Durchsetzung ihrer Normen zu erwarten ist. Zu diesem Zweck ist zunächst zu untersuchen, welcher Vorteil sich durch die Mitgliedschaft am Komplott, als der Übernahme der manipulierten Version V, ergibt.

Eine mögliche Ausrichtung von Komplotten besteht darin, durch gezielte Beeinflussung des Empfehlungssystems die Komplottmitglieder besser zu stellen, etwa um die Folgen ihres Betrugsverhaltens abzuschwächen. Hierfür ist gegenseitiges Hochloben der Komplottmitglieder vonnöten. In Abschnitt 12.2.3 haben wir jedoch bereits gesehen, dass im eigenen Ansatz der verteilten Vertrauensbildung durch das Empfehlungsmodell und die Verfahren der Glaubensbildung gegenseitiges Hochloben ineffektiv wird. Folglich kann der Sinn der Komplottbildung nicht in der Beeinflussung des Empfehlungssystems liegen.

Somit bleibt als einzig mögliche Ausrichtung eines Komplotts die Koordination des Verhaltens der Komplottmitglieder. In ihrem Verhalten diskriminieren sie Mitglieder von Nichtmitgliedern. Im Vergleich zur Gegenstrategie DiscriminatoryDefector aus Abschnitt 10.1.1 ergibt sich ein Vorteil darin, dass die Komplottmitglieder dieselben Kriterien für ihr diskriminierendes Verhalten anlegen. Diese Koordination des Verhaltens setzt allerdings die Abgrenzbarkeit von Mitgliedern zu Nichtmitgliedern voraus, die gemäß der obigen Betrachtungen nur zum Teil gegeben ist. Insgesamt ist die Vorteilhaftigkeit gegenüber der Gegenstrategie DiscriminatoryDefector somit nur sehr bedingt gegeben. Auf der anderen Seite führt der zusätzliche Einsatz der sozialen Kontrolle der Mitglieder untereinander zu einem nicht vernachlässigbaren Aufwand. Es ist somit fraglich, ob sich ein manipulationswilliger Benutzer nicht eher für die Gegenstrategie DiscriminatoryDefector entscheidet.

Ein weiteres Problem für die Herausbildung von Komplotten ist das opportunistische Verhalten unechter Komplottmitglieder. Dieses Problem wird besonders deutlich, wenn wir die Sicht eines manipulationswilligen Benutzers annehmen, der vor der Wahl zwischen den Versionen V und V' steht. Die Version V sieht vor, dass seine Einheit ein Mitglied des Komplotts wird und als solches die Normen des Komplotts befolgt. Diese Normen sehen eher kooperatives Verhalten gegenüber anderen Komplottmitgliedern vor. Durch den Verzicht auf Betrugsverhalten, entstehen somit für den Benutzer Opportunitätskosten, wenn er sich für die Version V entscheidet. Er muss diese Kosten nicht tragen, wenn er stattdessen die Version V' wählt. Im letzteren Fall gibt sich seine Einheit zwar als Komplottmitglied aus, sie bricht jedoch die Normen des Komplotts und teilt somit nicht die Opportunitätskosten der echten Komplottmitglieder. Aus diesen Überlegungen wird deutlich, dass sich ein manipulationswilliger Benutzer eher für die Version V' entscheidet. Damit besteht das Komplott aus vorwiegend unechten Komplottmitgliedern. In Analogie zu den Betrachtungen aus Abschnitt 5.3.2 führt dies zur Degeneration des Komplotts und letztendlich zum Aufhören seines Bestehens.

Berücksichtigung Normen-setzender Manipulation in der Evaluation. Die bisherigen Überlegungen deuten an, dass sich Komplotte im Informationssystem langfristig nicht durchsetzen können. Präzise Aussagen hierüber können allerdings nur für konkrete Ausgestaltungen von Komplotten getroffen werden. Im Zuge der Evaluation müssen somit viel versprechende Arten der Komplottbildung antizipiert und berücksichtigt werden. Mit der erforderlichen Erweiterung der Evaluationsmethodik beschäftigen wir uns im Folgenden.

In der Evaluationsmethodik aus Kapitel 9 wird zur Findung von Gegenstrategien ein eigenes Simulationswerkzeug, das Interaktive Kooperationsturnier, vorgeschlagen. In diesem durchforschen Versuchspersonen auf spielerische Weise den Raum der Gegenstrategien und zeigen dadurch viel versprechende Richtungen der Manipulation auf. Dieses Vorgehen ist notwendig, weil der Entwurfsraum der Gegenstrategien zu groß ist, als dass er erschöpfend untersucht werden könnte. Dieser Sachverhalt trifft umso mehr auf die Ausgestaltung von Komplotten zu, da der Spielraum hierfür gemäß der obigen Betrachtungen sehr viel größer ist. Daraus wird ersichtlich, dass auch der Entwurfsraum für Komplotte nicht erschöpfend untersucht werden kann.

Um dennoch auf systematische Weise viel versprechende Ausgestaltungen von Komplotten finden zu können, bietet sich eine Erweiterung des Interaktiven Kooperationsturniers an. Dazu müssen wir klären, wie Versuchspersonen auf spielerische Weise nicht Gegenstrategien sondern Ausgestaltungen von Komplotten identifizieren können. Hierfür ist eine grundsätzliche Änderung am Modell der Simulation notwendig: Im bisherigen Interaktiven Kooperationsturnier ist jede Versuchsperson für die Steuerung genau einer strategischen Einheit verantwortlich. Eine Versuchsperson kann jedoch nur dann ein Komplott entwerfen, wenn sie in der Simulation die Rolle eines Benutzers annimmt, der eine entsprechende manipulierte Version der Systemsoftware erstellt. Konkret bedeutet dies, dass jede Versuchsperson darüber bestimmt, wie sich eine Einheit verhält, die ihre manipulierte Version verwendet. Daraus wird deutlich, dass die Versuchspersonen nicht mehr länger das Verhalten jeweils einer Einheit steuern. Vielmehr steuern sie durch ihre Vorgaben jeweils das Verhalten von mehreren Einheiten.

Eine direkte Folge dieser Erweiterung des Interaktiven Kooperationsturniers ist, dass die Anforderungen an die teilnehmenden Versuchspersonen steigen. Diese müssen nunmehr in der Lage sein, auf deklarative oder imperative Weise zu bestimmen, wie sich die Einheiten verhalten sollen, die ihre Version übernehmen. Eine weitere zu lösende Fragestellung ist, nach welchen Kriterien im Interaktiven Kooperationsturnier bestimmt wird, wie viele strategische Einheiten von einer bestimmten manipulierten Version der Systemsoftware gesteuert werden. Um eine realisti-

sche Simulation zu ermöglichen, ist hierfür das Abschneiden der einzelnen Komplottmitglieder zu berücksichtigen. Dies liegt daran, dass im Falle einer erfolgreichen Komplottbildung weitere manipulationswillige Benutzer sich dazu entscheiden, mit ihrer Einheit am Komplott teilnehmen. In diesem Zusammenhang muss auch geklärt werden, wie eine Versuchsperson ihre Einheiten anweisen kann, sich als Mitglieder fremder Komplotte auszugeben.

Die Antwort auf diese Fragestellungen und die Umsetzung in eine erweiterte Evaluationsmethodik muss in zukünftigen Forschungsarbeiten erfolgen. Die in diesem Abschnitt angestellten Überlegungen bilden den Ausgangspunkt für solche Arbeiten.

12.4 Zusammenfassung

In diesem Kapitel wurden weiterführende Konzepte vorgestellt, die über den eigenen Ansatz hinausgehen. Dabei wurde zunächst die Fragestellung untersucht, ob der eigene Ansatz auch dann einsetzbar ist, wenn die Rahmenbedingungen, in denen sich das Informationssystem bewegt, im Vergleich zum Campus-Szenario erschwert sind.

Der Fokus dieser Untersuchung lag auf *einseitig vorteilhaften Transaktionen*. Es zeigt sich, dass das Zustandekommen solcher Transaktion notwendig ist, um Verzögerungen beim Erhalt benötigter Informationen zu vermeiden. Zu diesem Zweck wurde der Einsatz von Versprechen als prinzipieller Ansatz vorgestellt. Zwar kann aufgrund der Autonomie der Einheiten nicht erzwungen werden, dass Versprechen tatsächlich eingehalten werden. Die soziale Kontrolle zwischen den Einheiten im Zuge der verteilten Vertrauensbildung führt jedoch dazu, dass die Nichteinhaltung von Versprechen Folgekosten verursacht. Diese sind relativ beschränkt, wenn Versprechen abstreitbar sind. Daher wurden zwei Konzepte eingeführt, die auf die Nichtabstreitbarkeit von Versprechen beruhen. Ein Eigenwechsel beinhaltet ein Versprechen, das an den jeweiligen Gläubiger gebunden ist. Der Austausch der Eigenwechsel und ihr Einsatz im Rahmen des Empfehlungssystems und in der Glaubensbildung wurden besprochen. Inhaberwechsel gehen aus Eigenwechseln hervor, wenn das Anrecht auf die Einlösung eines Wechselversprechens übertragbar wird. Wir haben gezeigt, welche Vorteile sich daraus ergeben. Für beide Arten von nicht-abstreitbaren Versprechen wurden die offenen Fragestellungen herausgestellt, die für ihren Einsatz in zukünftigen Forschungsarbeiten zu beantworten sind. Abschließend wurde eine Übersicht der Anreize zum Eingehen von Transaktionen gegeben. Die dabei identifizierten Anreizmuster wurden untereinander in Beziehung gebracht.

Es wurden weitere Arten von Erschwernissen bezüglich der Rahmenbedingungen des Informationssystems besprochen. *Ungleiche Bedürfnisse und Fähigkeiten* führen dazu, dass der Grad der Kooperation im Gesamtsystem Einbußen erleidet. Als Substitut für fehlende Gegenleistungen im Informationssystem wurden daher realweltliche Aktionen eingeführt. Eine Folge daraus ist, dass Transaktionen zum Teil auf der realweltlichen Ebene und im Informationssystem stattfinden. Durch dieses Konzept kommt Kooperation immer dann zustande, wenn sie aus einer gesamtheitlichen Sicht sinnvoll ist. Außerdem sind wir auf *mehrseitige Transaktionen* eingegangen. Bei zusammengesetzten Aktionen übernimmt einer der Transaktionspartner die Rolle eines Entrepreneurs, der die Aktionen der Subkontraktoren zu einem Gesamtergebnis kombiniert. Wir haben gezeigt, dass hierfür eine Erweiterung der verteilten Vertrauensbildung wünschenswert aber nicht notwendig ist. Zusätzlich haben wir anhand des Lanes-Overlays untersucht, ob und wie die Robustheit von Kooperationsprotokollen gegenüber Fehlverhalten nachträglich erbracht werden kann. Weiterhin haben wir uns mit den Problemen beschäftigt, die sich aus *änderbaren Identitäten* ergeben. Das Problem gegenseitigen Hochlobens wird im eigenen Ansatz durch eine entsprechen-

de Bewertung von Bürgschaftsbeziehungen gelöst. Schwerwiegender ist das Problem, dass die Offenheit des Systems durch Möglichkeit zur Identitätsänderung gefährdet wird. Es wurden die Aufgaben aufgezeigt, die sich daraus für den Systementwerfer ergeben.

Abschließend wurden Erweiterungen des Entwurfs und der Evaluation der verteilten Vertrauensbildung aufgezeigt. Hierzu wurden zwei *alternative Transaktionsprotokolle* identifiziert, die weniger Aufwand für die Teilnehmer der Transaktion verursachen. Im Gegensatz zum asymmetrischen Transaktionsprotokoll erfordert das Vier-Wege Protokoll keine Erweiterungen am eigenen Ansatz der verteilten Vertrauensbildung. Anschließend wurde die Herausbildung von Komplotten untersucht, die vom Systementwurf *abweichende Normen* einsetzen. Hierzu wurde der Entwurfsraum für Komplotte analysiert und gezeigt, dass die Durchsetzung von Komplotten im Gesamtsystem nicht zu erwarten ist. Es wurde eine Erweiterung der Evaluationsmethodik vorgeschlagen, mit Hilfe derer diese Vermutung in zukünftigen Arbeiten überprüft werden kann.

Kapitel 13

其言之不作 则其為之也難

"Sei nicht bereit, über eine Sache zu sprechen, außer wenn ihre Umsetzung nicht in deiner Macht liegt."
(Gespräche und Aussprüche des Konfuzius, 14.21)

Ausblick

In diesem abschließenden Kapitel erörtern wir, welche Auswirkungen diese Dissertation auf zukünftige Forschungs- und Entwicklungsarbeiten hat. Dabei werden wir von der Frage geleitet, in welche Richtungen eine Weiterentwicklung des eigenen Ansatzes sinnvoll oder (für den Einsatz in anderen Szenarien) gar notwendig ist.

In dieser Arbeit konnten wir die These von der Existenzfähigkeit von Informationssystemen, die wie im Campus-Szenario vollständig auf die Geräte autonomer Benutzer verteilt sind, validieren. Es liegt daher nahe, ein solches Informationssystem für das universitäre Umfeld zu entwickeln und anzuwenden. Die dazu notwendige Erschaffung der originalen Systemsoftware stellt zwar hauptsächlich Entwicklungsarbeit dar, die auf die prototypische Implementierung des eigenen Ansatzes im Rahmen des Simulativen Kooperationsturniers aufbauen kann. Der Einsatz im realen Umfeld wirft jedoch auch einige Forschungsfragestellungen auf. Von besonderer Bedeutung ist hierbei die Gestaltung der Schnittstelle zwischen dem Benutzer und seiner Einheit. Diese könnte nicht nur für die Kommunikation der Benutzerwünsche und der Darstellung der erhaltenen Informationen sorgen. Zudem scheint es sinnvoll, dass der Benutzer über einige Vorgänge innerhalb des Informationssystems informiert wird. Ein Beispiel hierfür ist eine Statistik darüber, in wie vielen Transaktionen seine Einheit betrogen worden ist. Diese Transparenz erhöht das Vertrauen der Benutzer in das Informationssystem und hilft ihnen, die Güte der originalen Systemsoftware zu beurteilen. Für die Darstellung der Informationssystem-internen Vorgänge könnte die Benutzungsschnittstelle des Interaktiven Kooperationsturniers als Ausgangspunkt dienen. Diese sorgt für die visuelle Darstellung aller Vorgänge, die aus der Sicht der Versuchspersonen relevant ist.

Der prototypische Einsatz des Informationssystems im realen Umfeld würde zudem Rückschlüsse für die Gestaltung der Simulationswerkzeuge erlauben. Von besonderem Interesse sind hierbei die soziologischen und psychologischen Faktoren, die die Benutzer in ihrer Wahl der verwendeten Systemsoftware beeinflussen. Durch entsprechende Studien ließe sich von den Benutzern in der Simulation ein Bild zeichnen, das differenzierter als das rein utilitaristische Modell dieser Arbeit ist.

Eine Reihe weiterführender Forschungsarbeiten wird benötigt, um den eigenen Ansatz auch in Umgebungen einsetzen zu können, deren Charakteristika sich vom Campus-Szenario auf ent-

scheidende Weise unterscheiden. Im Folgenden gehen wir auf verschiedene Charakteristika ein, die eine Erweiterung des eigenen Ansatzes notwendig oder zumindest wünschenswert machen.

Im Campus-Szenario ist die Menge der Benutzer relativ homogen. Dies liegt daran, dass es sich bei den Benutzern um Studenten handelt, die für den Zeitraum ihres Studiums am Informationssystem teilnehmen und dabei ähnliche Ziele verfolgen. Ein anderes Bild ergibt sich jedoch, wenn die Menge der Benutzer *heterogen* ist. Ein Beispiel hierfür ist ein auswärtiger Besucher der Universität, der sich nur für einen Tag in der Universität und ihrer Umgebung aufhält. Eine etwaige Teilnahme am Informationssystem wäre für diesen Besucher von sehr kurzer Dauer. Seine Einheit wird somit mit dem Problem konfrontiert, dass sie als Neuankömmling im sozialen Gefüge nicht dieselbe Stellung wie die länger ansässigen Einheiten einnehmen kann. Dies führt unter Umständen dazu, dass die Einheit des Besuchers nicht als Transaktionspartner angenommen wird und der Besucher somit faktisch aus dem Informationssystem ausgeschlossen wird. Als Ausgangspunkt für die Bearbeitung dieses Problems bieten sich die weiterführenden Konzepte aus Kapitel 12.2 an. Der Besucher könnte über seinen Kontakt mit einem weiteren Benutzer, dessen Einheit im sozialen Gefüge etabliert ist, in Besitz von Inhaberwechseln und Informationen über die Vertrauenswürdigkeit anderer Einheiten gelangen. Wenn es sich beim Besucher zum Beispiel um einen Gastvortragenden handelt, so bietet sich zu diesem Zweck sein Kontakt zum einladenden Universitätsangestellten an. Weiterhin erhalten soziale Bindungen zwischen Einheiten durch die Heterogenisierung der Benutzermenge eine größere Bedeutung, so dass sich unter Umständen außer der beidseitigen Bürgschaft weitere Formen der Beziehungen anbieten.

Eine weitere Richtung zukünftiger Forschungsarbeiten ergibt sich für Szenarien, in denen das Informationssystem durch die Verwendung vertrauenswürdiger Dritter nur zum Teil selbstorganisierend ist. Der hierfür erforderliche Einsatz eines ausfallsicheren und robusten *Servers* und die Sicherstellung seiner ubiquitären Erreichbarkeit würden dem Betreiber des Informationssystems zwar einen erheblichen Aufwand verursachen. Dieser ist jedoch in kommerziellen Szenarien, in denen die Teilnahme am Informationssystem kostenpflichtig ist, unter Umständen durchaus vertretbar. Um die Anforderungen an den Betrieb und die Erreichbarkeit eines zentralen Servers in Grenzen zu erhalten, bietet es sich an, dass ihm lediglich eine unterstützende Funktion bei der Glaubensbildung der einzelnen Einheiten zukommt. Konkret bedeutet dies, dass der Server eine zentrale Ablage von Beweismitteln und Empfehlungen wäre. Dies hätte darüber hinaus zum Vorteil, dass dem Betreiber des Servers nicht voll vertraut werden muss, da er die Bestandteile der zentralen Ablage (also die Beweismittel) nicht manipulieren sondern lediglich selektiv entfernen kann.

Weiterhin ist zu untersuchen, welche Auswirkungen sich ergeben, wenn die Menge der Benutzer sehr viel weiter als im Campus-Szenario gefasst wird. Dies ist dann der Fall, wenn das Informationssystem nicht nur für die Studenten einer Universität sondern für die Bürger einer Stadt oder gar eines ganzen Landes ausgelegt ist. In einem solchen *Bürgernetz* würde die Zahl der Einheiten 100.000 oder mehr betragen. Dieses Szenario wird zum Beispiel im BMBF-Forschungsschwerpunkt Internetökonomie ins Auge gefasst [BMB05]. Bei dieser Systemgröße sind die in dieser Arbeit vorgeschlagenen Verfahren zur verteilten Vertrauensbildung anders auszurichten. So ist zu erwarten, dass nicht einzelne Transaktionserfahrungen sondern soziale Bindungen die Grundlage für die soziale Kontrolle zwischen den Einheiten bilden. Ob und unter welchen Umständen der Einsatz transaktionaler Beweismittel bei solchen Systemgrößen sinnvoll ist, bleibt eine offene Fragestellung. Des Weiteren werden Inhaberwechsel in einem solchen Bürgernetz wohl eine größere Rolle als Eigenwechsel und direkte Gegenleistungen spielen, da Wiederbegegnungen in der Regel weniger wahrscheinlich sind. Außerdem ist in zukünftigen Forschungsarbeiten über Bürgernetze zu untersuchen, wie sich der organisatorische Zusammenschluss einer Menge von Benutzern (etwa

den Angestellten eines Unternehmens) auf die Effektivität des eigenen Ansatzes auswirkt. Eine solche Benutzergruppe könnte gemeinschaftliche Ziele im Informationssystem verfolgen, da ihre Mitglieder untereinander gebunden und somit nicht autonom sind.

Der Ansatz dieser Dissertation und der zukünftigen Forschungsarbeiten, die ihn weiterführen, stehen in einem größeren Gesamtzusammenhang. Im Mittelpunkt stehen *künstliche Gesellschaften*, die aus den Einheiten autonomer Benutzer bestehen, und die Verfahren, die für soziale Kontrolle zwischen den Einheiten sorgen und damit die Herausbildung eines sozialen Gefüges zur Folge haben. Die Verfahren dieser Arbeit sind am Campus-Szenario ausgerichtet und führen aufgrund seiner Eigenschaften zu künstlichen Gesellschaften, die sowohl egalitär als auch von ihrer Größe her überschaubar sind. Die zukünftigen Forschungsarbeiten werden zu künstlichen Gesellschaften führen, in denen Einheiten bestimmte soziale Rollen einnehmen und deren Größe derjenigen der menschlichen Gesellschaft gleichkommt. Insofern nähert sich die künstliche der menschlichen Gesellschaft in Größe und Struktur an. Gleichzeitig haben die Verfahren dieser Dissertation und der zukünftigen Arbeiten zum Ziel, dass sich das soziale Gefüge in der künstlichen Gesellschaft automatisiert, also ohne Zutun der jeweiligen Benutzer, herausbildet. Dies führt dazu, dass die Rollen, die ein Benutzer und seine Einheit in der menschlichen beziehungsweise in der künstlichen Gesellschaft einnehmen, grundsätzlich verschieden sein können. Es bleibt daher letztlich für die Forschung die Frage, ob und wie sich ein Zusammenhang zwischen den an sich nebeneinander existierenden künstlichen und menschlichen Gesellschaften herstellen lässt.

Literaturverzeichnis

[ABKM01] Andersen, David G., Hari Balakrishnan, M. Frans Kaashoek und Robert Morris: *Resilient Overlay Networks*. In: *Proceedings of the 18th ACM Symposium on Operating Systems Principles (SOSP)*, Banff, Kanada, Oktober 2001.

[AFS97] Arbaugh, Bill, Dave Farber und Jonathan Smith: *A Secure and Reliable Bootstrap Architecture*. In: *Proceedings of the IEEE Symposium on Security and Privacy*, Seiten 65–71, 1997.

[AK96] Anderson, Ross und Markus Kuhn: *Tamper Resistance - a Cautionary Note*. In: *Proceedings of the Second Usenix Workshop on Electronic Commerce*, Seiten 1–11, 1996.

[Ake70] Akerlof, George A.: *The Market for Lemons: Quality Uncertainty and the Market Mechanism*. Quarterly Journal of Economics, 89:488–500, 1970.

[And03a] Anders, Ralf: *Anreizmuster und Ihre Einsatzmöglichkeiten in Mobilen Ad-Hoc-Netzen*, 2003. Diplomarbeit an der Fakultät für Informatik, Universität Karlsruhe.

[And03b] Anderson, Ross: *Cryptography and Competition Policy - Issues with Trusted Computing*. In: *2nd Annual Workshop on Economics and Information Security*, Maryland, 2003.

[And03c] Anderson, Ross: *TCPA/Palladium FAQ*. im Internet verfügbar unter http://www.cl.cam.ac.uk/users/rja14/tcpa-faq.html, 2003.

[ANL00] Aura, Tuomas, Pekka Nikander und Jussipekka Leiwo: *DOS-Resistant Authentication with Client Puzzles*. Proceedings of the Security Protocols Workshop, LNCS, 2133:170–177, 2000.

[AO03] Anders, Ralf und Philipp Obreiter: *Economic Incentive Patterns and their Application to Ad Hoc Networks*. Technischer Bericht 2003-17, Universität Karlsruhe, Fakultät der Informatik, Oktober 2003.

[ARH97] Abdul-Rahman, Alfarez und Stephen Hailes: *A Distributed Trust Model*. In: *Proceedings of the ACM New Security Paradigms Workshop '97*, Seiten 48–60, Great Langdale, UK, September 1997.

[Aso98] Asokan, Nadarajah: *Fairness in Electronic Commerce*. Doktorarbeit, University of Waterloo, 1998.

[ASST96] Ashlock, Dan, Mark D. Smucker, E. Ann Stanley und Leigh Tesfatsion: *Preferential Partner Selection in an Evolutionary Study of Prisoner's Dilemma.* BioSystems, 37(1):99–125, 1996.

[Axe84] Axelrod, Robert: *The Evolution of Cooperation.* Basic Books, 1984.

[Axe97] Axelrod, Robert: *Advancing the Art of Simulation in the Social Sciences.* In: Conte, R., R. Hegselmann und P. Terna (Herausgeber): *Simulating Social Phenomena*. Springer-Verlag, 1997.

[Axe00] Axelrod, Robert: *On Six Advances in Cooperation Theory.* Analyse & Kritik, 22:130–151, 2000.

[Bac90] Bacchus, Fahiem: *Probabilistic Belief Logics.* In: *Proceedings of European Conference on Artificial Intelligence (ECAI-90)*, Seiten 59–64, 1990.

[BB02a] Buchegger, Sonja und Jean-Yves Le Boudec: *Nodes Bearing Grudges: Towards Routing Security, Fairness, and Robustness in Mobile Ad Hoc Networks.* In: *Proc. of the Tenth Euromicro Workshop on Parallel, Distributed and Network-based Processing*, Seiten 403 – 410, Canary Islands, Spanien, Januar 2002. IEEE Computer Society.

[BB02b] Buchegger, Sonja und Jean-Yves Le Boudec: *Performance Analysis of the CONFIDANT Protocol: Cooperation Of Nodes — Fairness In Distributed Ad-hoc NeTworks.* In: *Proc. of IEEE/ACM Workshop on Mobile Ad Hoc Networking and Computing (MobiHOC)*, Seiten 226–236, Lausanne, Schweiz, Juni 2002.

[BB03] Buchegger, Sonja und Jean-Yves Le Boudec: *The Effect of Rumor Spreading in Reputation Systems for Mobile Ad-hoc Networks.* In: *Proceedings of WiOpt '03: Modeling and Optimization in Mobile, Ad Hoc and Wireless Networks*, Sophia-Antipolis, Frankreich, März 2003.

[BB04a] Buchegger, Sonja und Jean-Yves Le Boudec: *A Robust Reputation System for P2P and Mobile Ad-Hoc Networks.* In: *Second Workshop on the Economics of Peer-to-Peer Systems*, Harvard, MA, 2004.

[BB04b] Buchmann, Erik und Klemens Böhm: *FairNet - How to Counter Free Riding in Peer-to-Peer Data Structures.* In: *Proc. of the International Conference on Cooperative Information Systems 2004, Agia Napa, Zypern*, 2004.

[BB05] Böhm, Klemens und Erik Buchmann: *Free Riding-Aware Forwarding in Content-Addressable Networks.* The VLDB Journal, 1, 2005. Im Erscheinungsprozess.

[BCL+04] Bhargava, Bharat, Elizabeth Chang, Leszek Lilien, Farookh Khadeer Hussain, Arnon Rosenthal, Wolfgang Nejdl, Marianne Winslett, Daniel Olmedilla, Morris Sloman, Vipul Kashyap und Tharam Dillon: *The Pudding of Trust.* IEEE Intelligent Systems, 19(5):74–88, 2004.

[Ber85] Berger, James: *Statistical Decision Theory and Bayesian Analysis.* Springer Series in Statistics, 1985.

[BG97] Bacharach, Michael und Diego Gambetta: *Trust in Signs.* In: Cook, K.S. (Herausgeber): *Trust in Society*, Seiten 148–184. Russell Sage Foundation. New York, 1997.

[BH03] Buttyan, Levente und Jean-Pierre Hubaux: *Stimulating Cooperation in Self-Organizing Mobile Ad Hoc Networks.* ACM/Kluwer Mobile Networks and Applications (MONET), 8(5), 2003.

[BJHS03] Buttyan, Levente, Markus Jakobsson, Jean-Pierre Hubaux und Naouel Ben Salem: *Incentive Mechanisms in Multi-Hop Wireless Networks.* In: *Proc. of WiOpt 03: Modeling and Optimization in Mobile, Ad Hoc and Wireless Networks (WiOpt'03)*, Sophia-Antipolis, Frankreich, 2003.

[BKOKR04] Breyer, Tobias, Michael Klein, Philipp Obreiter und Birgitta König-Ries: *Activity-Based User Modeling in Service-Oriented Ad-Hoc-Networks.* In: *First Working Conference on Wireless On-Demand Network Systems (WONS 2004)*, Trento, Italien, Januar 2004.

[BMB05] BMBF-Forschungsschwerpunkt Internetökonomie, Universität Karlsruhe: *Selbstorganisation und Spontaneität in liberalisierten und harmonisierten Märkten.* http://www.internetoekonomie.uni-karlsruhe.de, 2005.

[Bra01] Brainov, Sviatoslav: *An Incentive Compatible Trading Mechanism for Trust Revelation.* In: *In Proceeding of the IJCAI'01 Workshop Economic Agents, Models and Mechanisms*, Seiten 62–70, 2001.

[BS89] Bamberg, Guenter und Klaus Spremann: *Agency Theory, Information, and Incentives.* Springer, 1989.

[BS99] Brainov, Sviatoslav und Tuomas Sandholm: *Contracting with Uncertain Level of Trust.* In: *Proceedings of the First ACM Conference on Electronic Commerce*, Seiten 15–21, Denver, USA, 1999.

[BSGA01] Buyya, Rajkumar, Heinz Stockinger, Jonathan Giddy und David Abramson: *Economic Models for Management of Resources in Peer-to-Peer and Grid Computing.* In: *Technical Track on Commercial Applications for High-Performance Computing, SPIE International Symposium on The Convergence of Information Technologies and Communications (ITCom'01)*, Denver, USA, 2001.

[But01] Buttyan, Levente: *Building Blocks for Secure Services: Authenticated Key Transport and Rational Exchange Protocols.* Doktorarbeit, EPFL, 2001.

[CA02] Courcoubetis, Costas und Panayotis Antoniadis: *Market Models for P2P Content Distribution.* In: *Proc. of the First Intl. Workshop on Agents and Peer-To-Peer Computing (AP2PC)*, Bologna, Italien, 2002.

[CCP98] Castelfranchi, Cristiano, Rosaria Conte und Mario Paolucci: *Normative Reputation and the Costs of Compliance.* Journal of Artificial Societies and Social Simulation, 1(3), 1998.

[Cha98] Chattoe, Edmund: *Just How (Un)Realistic are Evolutionary Algorithms as Representations of Social Processes?* Journal of Artificial Societies and Social Simulation, 1(3), 1998.

[Cou91] Council of the European Communities: *Software Directive – Council Directive on the Legal Protection of Computer Programs (91/250/EEC)*, Mai 1991.

[Cou04] Council of the European Communities: *Council Directive on the Enforcement of Intellectual Property Rights (2004/48/EC)*, April 2004.

[CP02] Conte, Rosaria und Mario Paolucci: *Reputation in Artificial Societies. Social Beliefs for Social Order.* Kluwer, Boston, 2002.

[CWJS97] Crow, Brian P., Indra Kim Widjaja, Geun Jeong und Prescott T. Sakai: *IEEE 802.11 Wireless Local Area Networks.* IEEE Communications Magazine, Seiten 116–126, September 1997.

[DA04] Despotovic, Zoran und Karl Aberer: *A Probabilistic Approach to Predict Peers' Performance in P2P Networks.* In: *8th Intl Workshop on Cooperative Information Agents (CIA'04)*, Erfurt, Deutschland, 2004.

[Del00] Dellarocas, Chrysanthos: *Immunizing online reputation reporting systems against unfair ratings and discriminatory behavior.* In: *Proceedings of the ACM Conference on Electronic Commerce*, Seiten 150–157, Minneapolis, MN, USA, 2000.

[Del03] Dellarocas, Chrysanthos: *The Digitization of Word-of-Mouth: Promise and Challenges of Online Feedback Mechanisms.* Management Science, 2003.

[Dör04] Dörr, Claudia: *Entwicklung einer Komponente zur Beweismittelverwaltung und Beweisführung*, April 2004. Studienarbeit an der Fakultät für Informatik, Universität Karlsruhe.

[DUK00] DUKATH: *Wireless Internet Access at the University of Karlsruhe.*, 2000. http://www.uni-karlsruhe.de/Uni/RZ/ Netze/DUKATH/dukath-engl.html.

[eBa03] eBay Inc.: *AuctionWeb Server*, 2003. http://www.ebay.com/.

[ECCC94] Elzer, Stephanie, Jennifer Chu-Carroll und Sandra Carberry: *Recognizing and Utilizing User Preferences in Collaborative Consultation Dialogues.* In: *Proceedings of the Fourth International Conference on User Modeling*, Seiten 19–24, 1994.

[Fäh05] Fähnrich, Stefan: *Evaluation normativer Strategien des verteilten Reputationssystem EVIDIRS*, Mai 2005. Diplomarbeit an der Fakultät für Informatik, Universität Karlsruhe.

[FH03] Friedman-Hill, Ernest: *Jess in Action, Rule-Based Systems in Java.* Manning, 2003.

[FKÖD04] Fernandes, Alberto, Evangelos Kotsovinos, Sven Östring und Boris Dragovic: *Pinocchio: Incentives for Honest Participation in Distributed Trust Management.* In: *Second International Conference on Trust Management (iTrust'04)*, Seiten 63–77, Oxford, UK, 2004. Springer LNCS 2995.

[FO04] Fähnrich, Stefan und Philipp Obreiter: *The Buddy System - A Distributed Reputation System Based On Social Structure.* Technischer Bericht 2004-1, Universität Karlsruhe, Fakultät der Informatik, Februar 2004.

[FOKR04] Fähnrich, Stefan, Philipp Obreiter und Birgitta König-Ries: *The Buddy System: A Distributed Reputation System Based on Social Structure.* In: *7th Intl. Workshop on Data Management in Mobile Environments*, Ulm, September 2004.

[FT91] Fudenberg, Drew und Jean Tirole: *Game Theory.* MIT Press, Cambridge, Massachusetts, 1991.

[GLBML01] Golle, Philippe, Kevin Leyton-Brown, Ilya Mironov und Mark Lillibridge: *Incentives for Sharing in Peer-to-Peer Networks.* In: *Proceedings of the ACM Conference on Electronic Commerce*, Band 2232, Seiten 75–86, Tampa, FL, USA, 2001.

[Gri03] Gries, Matthias: *Methods for Evaluating and Covering the Design Space During Early Design Development.* Technischer Bericht UCB/ERL M03/32, Electronics Research Lab, University of California at Berkeley, August 2003.

[Hei02] Heise online: *Experten Warnen Vor Massiven Problemen Bei TCPA und Palladium.* http://www.heise.de/newsticker/meldung/33330, Dezember 2002.

[Hel97] Helton, Jon C.: *Uncertainty and Sensitivity Analysis in the Presence of Stochastic and Subjective Uncertainty.* Journal of Statistical Computation and Simulation, 57(1–4):3–76, 1997.

[Hey04] Heylighen, Francis: *The Science of Self-Organization and Adaptivity.* Encyclopedia of Life Support Systems, 2004. http://www.eolss.net/.

[HM84] Halpern, Joseph Y. und Yoram Moses: *Knowledge and Common Knowledge in a Distributed Environment.* In: *Symposium on Principles of Distributed Computing*, Seiten 50–61, 1984.

[Hof00] Hoffmann, Robert: *Twenty Years on: The Evolution of Cooperation Revisited.* Journal of Artificial Societies and Social Simulation, 2000.

[Hof02] Hoffman, Ivan: *Derivative Works*, 2002. Im Internet verfügbar unter http://www.ivanhoffman.com/derivative2.html.

[ISO97] ISO: *ISO/IEC 13888*, 1997.

[JF03] Jurca, Radu und Boi Faltings: *Towards Incentive-Compatible Reputation Management.* In: al., R. Falcone et (Herausgeber): *AAMAS'02-Workshop on Deception, Fraud and Trust in Agent Societies.* Springer LNAI 2631, 2003.

[JHB03] Jakobsson, Markus, Jean-Pierre Hubaux und Levente Buttyan: *A Micro-Payment Scheme Encouraging Collaboration in Multi-Hop Cellular Networks.* In: *Proc. of the Seventh International Financial Cryptography Conference*, Guadeloupe, Frankreich, 2003.

[JI02] Josang, Audun und Roslan Ismail: *The Beta Reputation System.* In: *15th Bled Conference on Electronic Commerce*, Bled, Slowenien, Juni 2002.

[JKSS04] Jondral, Friedrich, Clemens Klöck, Henrik Schober und Gunther Sessler: *Einbindung Von Campusnetzwerken in Die UMTS-Infrastruktur.* In: Deussen, Juling, Thum (Herausgeber): *Nukath - Die Notebook Universität Karlsruhe (TH)*, Seiten 130–141. Universitätsverlag Karlsruhe, 2004.

[Jon05] Jones, Pamela: *Software, Reverse Engineering and the Law*, Mai 2005. http://lwn.net/Articles/134642/.

[JT05] Johnson, Addie und Niels A. Taatgen: *The Handbook of Human Factors in Web Design*, Kapitel User Modeling, Seiten 424–438. Erlbaum, 2005.

[Kau93] Kauffman, Stuart A.: *The Origins of Order: Self-Organization and Selection in Evolution.* Oxford University Press, 1993.

[KaZ05] KaZaA: *End User License Agreement*, 2005. http://www.kazaa.com/us/terms2.htm.

[KDMT00] Kfir-Dahav, Noa E., Dov Monderer und Moshe Tennenholtz: *Mechanism design for resource bounded agents.* In: *Fourth International Conference on Multi-Agent Systems (ICMAS-2000)*, Barcelona, Spanien, 2000.

[Kir05] Kiratiwintakorn, Phongsak: *Energy Efficient Security Framework for Wireless Local Area Networks.* Doktorarbeit, University of Pittsburgh, 2005.

[KKRO03] Klein, Michael, Birgitta König-Ries und Philipp Obreiter: *Lanes – A Lightweight Overlay for Service Discovery in Mobile Ad Hoc Network.* In: *Proc. of the 3rd Workshop on Applications and Services in Wireless Networks (ASWN2003)*, Berne, Schweiz, Juli 2003.

[Kle03] Klein, Michael: *DIANEmu – A Java-Based Generic Simulation Environment for Distributed Protocols.* Technischer Bericht 2003-7, Universität Karlsruhe, Fakultät der Informatik, Mai 2003.

[KN03] Koenig, Christian und Andreas Neumann: *Anforderungen Des EG-Wettbewerbsrechts an Vertrauenswürdige Systemumgebungen TCPA, TCG, Palladium und NGSCB.* MultiMedia und Recht, 11:695–700, 2003.

[Kou03] Kou, Weidong: *Payment Technologies for E-Commerce.* Springer, 2003.

[KR03] Kinateder, Michael und Kurt Rothermel: *Architecture and Algorithms for a Distributed Reputation System.* In: Nixon, P. und S. Terzis (Herausgeber): *Proc. Of the First Intl. Conf. On Trust Management (iTrust)*, Seiten 1–16, Heraklion, Griechenland, 2003. Springer LNCS 2692.

[KRKB04] König-Ries, Birgitta, Michael Klein und Tobias Breyer: *Activity-Based User Modeling in Wireless Networks.* Mobile Networks and Applications. Special Issue on Internet Wireless Access: 802.11 and Beyond, 2004.

[KS92] Kreps, David und Joel Sobel: *Signalling.* In: Aumann, Robert und Sergiu Hart (Herausgeber): *Handbook of Game Theory*, Kapitel 25, Seiten 849–867. Elsevier Science Publishers, 1992. Volume 1.

[KSGM03] Kamvar, Sepandar D., Mario T. Schlosser und Hector Garcia-Molina: *The EigenTrust Algorithm for Reputation Management in P2P Networks.* In: *WWW2003*, Budapest, Ungarn, 2003.

[KW82] Kreps, David und Robert Wilson: *Reputation and Imperfect Information.* Journal of Economic Theory, 27:253–279, 1982.

[LD03] Linn, Cullen und Saumya Debray: *Obfuscation of Executable Code to Improve Resistance to Static Disassembly.* In: *Proceedings of the 10th ACM Conference on Computer and Communication Security*, Seiten 290–299, 2003.

[LI04] Liu, Jinshan und Valerie Issarny: *Enhanced Reputation Mechanism for Mobile Ad Hoc Networks.* In: *Second International Conference on Trust Management (iTrust'04)*, Seiten 48–62, Oxford, UK, 2004. Springer LNCS 2995.

[LO03] Liu, Lei und Philipp Obreiter: *The Software Station - A System for Version Controlled Development and Web Based Deployment of Software for a Mobile Environment.* In: *Proc. of the National Data Base Conference (NDBC'03)*, Changsha, China, 2003.

[LSP82] Lamport, Leslie, Robert Shostak und Marshall Pease: *The Byzantine Generals Problem.* ACM Transactions of Programming Languages and Systems, 4(3):382–401, Juli 1982.

[Mar94] Marsh, Steve: *Formalising Trust as a Computational Concept.* Doktorarbeit, Department of Mathematics and Computer Science, University of Stirling, 1994.

[MGLB00] Marti, Sergio, T. J. Giuli, Kevin Lai und Mary Baker: *Mitigating routing misbehavior in mobile ad hoc networks.* In: *Mobile Computing and Networking*, Seiten 255–265, 2000.

[MGM04] Marti, Sergio und Hector Garcia-Molina: *Limited Reputation Sharing in P2P Systems.* In: *ACM Conference on Electronic Commerce (EC'04)*, New York, Mai 2004.

[MHM02] Mui, Lik, Ari Halberstadt und Mojdeh Mohtashemi: *Notions of Reputation in Multi-Agents Systems: A Review.* In: *Proceedings of the First International Joint Conference on Autonomous Agents and Multiagent Systems (AAMAS'02)*, Bologna, Italien, 2002.

[MM02a] Michiardi, Pietro und Refik Molva: *Core: A Collaborative Reputation Mechanism To Enforce Node Cooperation In Mobile AD HOC Networks.* In: *Proceedings of the 6th IFIP Communications and Multimedia Security Conference, Portorosz, Slowenien*, 2002.

[MM02b] Michiardi, Pietro und Refik Molva: *Making greed work in mobile ad hoc networks.* Technischer Bericht, Institut Eurécom, 2002.

[MO04] Mahler, Tobias und Thomas Olsen: *Reputation Systems and Data Protection Law.* In: *Third Internal iTrust Workshop on Trust Management in Dynamic Open Systems*, Dortmund, Deutschland, 2004.

[MR82] Milgrom, Paul und J. Roberts: *Predation, Reputation, and Entry Deterrence.* Economic Theory, 27:280–312, 1982.

[MR03] Miranda, Hugo und Lus Rodrigues: *Friends and Foes: Preventing Selfishness in Open Mobile Ad Hoc Networks.* In: *Proc. of the First Intl. Workshop on Mobile Distributed Computing (MDC'03)*, Providence, RI, USA, 2003. IEEE Computer Society Press.

[MR04] Maier-Rigaud, Frank P.: *Externality or Common Good? Choosing an Adequate Framework to Analyze Institutional Aspects of Common Goods.* Preprint Series of the Max Planck Project Group on Common Goods, 2004.

[MS82] Maynard Smith, James: *Evolution and the Theory of Games.* Cambridge University Press, 1982.

[MT03] Moreton, Tim und Andrew Twigg: *Enforcing Collaboration in Peer-to-Peer Routing Services.* In: *Proceedings of the First Intl. Conf. on Trust Management (iTrust)*, Seiten 255–270, Heraklion, Crete, Griechenland, 2003.

[Mus98] Musker, David: *Reverse Engineering.* In: *Protecting and Exploiting Intellectual Property in Electronics (IBC Conferences)*, Juni 1998.

[ND04] Nordemann, Jan Bernd und Andreas Dustmann: *To Peer Or Not To Peer - Urheberrechtliche und Datenschutzrechtliche Fragen der Bekämpfung der Internet Piraterie.* Computer und Recht, Seite 380ff, Mai 2004.

[Nor98] Norman, Don A.: *The Invisible Computer.* MIT Press, Cambridge, MA, 1998.

[Obr04a] Obreiter, Philipp: *A Case for Evidence-Aware Distributed Reputation Systems.* In: *Second International Conference on Trust Management (iTrust'04)*, Seiten 33–47, Oxford, UK, 2004. Springer LNCS 2995.

[Obr04b] Obreiter, Philipp: *Cooperation Incentives.* In: *Dagstuhl Seminar No. 04441 on Mobile Information Management*, Oktober 2004.

[Obr05] Obreiter, Philipp: *Bewegliche Objekte in Datenbanken.* In: Höpfner, Hagen, Can Türker und Birgitta König-Ries (Herausgeber): *Mobile Datenbanken und Informationssysteme*, Kapitel 6, Seiten 105–124. dPunkt Verlag, Heidelberg, 2005.

[OFN04] Obreiter, Philipp, Stefan Fähnrich und Jens Nimis: *How Social Structure Improves Distributed Reputation Systems - Three Hypotheses.* In: *Third Intl. Workshop on Agents and Peer-to-Peer Computing (AP2PC'04), Springer LNCS 3601*, New York, 2004.

[OG02] Obreiter, P. und G. Graef: *Towards Scalability in Tuple Spaces.* In: *ACM Symposium of Applied Computing (SAC) Special Track on Coordination Models, Languages and Applications*, Seiten 344–350, Madrid, Spanien, 2002.

[OK03a] Obreiter, Philipp und Michael Klein: *Self-Configuring Resource Management in Cooperative and Uncooperative Autonomous Systems.* Technischer Bericht 2003-15, Universität Karlsruhe, 2003.

[OK03b] Obreiter, Philipp und Michael Klein: *Vertical Integration of Incentives for Cooperation - Inter-Layer Collaboration as a Prerequisite for Effectively Stimulating Cooperation in Ad Hoc Networks.* In: *Proc. of the Second Mediterranean Workshop on Ad-Hoc Networks (MED-HOC NET 2003)*, Mahdia, Tunesien, 2003.

[OKR05] Obreiter, Philipp und Birgitta König-Ries: *A New View on Normativeness in Distributed Reputation Systems – Beyond Behavioral Beliefs.* In: *Fourth Workshop on Agents and Peer-to-Peer Computing (AP2PC'05)*, Utrecht, Niederlande, 2005.

[OKRK03] Obreiter, Philipp, Birgitta König-Ries und Michael Klein: *Stimulating Cooperative Behavior of Autonomous Devices - An Analysis of Requirements and Existing Approaches.* In: *Proceedings of the Second International Workshop on Wireless Information Systems (WIS2003)*, Seiten 71–82, Angers, Frankreich, 2003.

[OKRP04] Obreiter, Philipp, Birgitta König-Ries und Georgios Papadopoulos: *Engineering Incentive Schemes for Ad Hoc Networks - A Case Study for the Lanes Overlay.* In: *First EDBT-Workshop on Pervasive Information Management, LNCS 3268*, 2004.

[ON03] Obreiter, Philipp und Jens Nimis: *A Taxonomy of Incentive Patterns - The Design Space of Incentives for Cooperation.* In: *Second Intl. Workshop on Agents and Peer-to-Peer Computing (AP2PC'03), Springer LNCS 2872*, Melbourne, Australien, 2003.

[ON04] Obreiter, Philipp und Ioana Nistoreanu: *Transaction Protocols for Self-Organizing Systems of Autonomous Entities.* Technischer Bericht 2004-11, Universität Karlsruhe, Fakultät der Informatik, Juli 2004.

[Pap03] Papadopoulos, Georgios: *S-Lanes - Ein schlankes Overlay zur motivierten Dienstvermittlung in mobilen Ad-Hoc-Netzen*, Dezember 2003. Diplomarbeit an der Fakultät für Informatik, Universität Karlsruhe.

[PS04] Poellabauer, Christian und Karsten Schwan: *Energy-Aware Media Transcoding in Wireless Systems.* In: *Proceedings of the 10th Real-Time and Embedded Technology and Applications Symposium (RTAS)*, 2004.

[PS05] Papaioannou, Thanasis G. und George D. Stamoulis: *Optimizing an Incentives' Mechanism for Truthful Feedback in Virtual Communities.* In: *Fourth International Workshop on Agents and Peer-to-Peer Computing (AP2PC)*, Utrecht, Niederlande, 2005.

[PSW98] Pfitzmann, Birgit, Matthias Schunter und Michael Waidner: *Optimal Efficiency of Optimistic Contract Signing.* In: *Proceedings of the 7th Annual ACM Symposium on Principles of Distributed Computing (PODC)*, Seiten 113–122, Puerto Vallarta, Mexico, 1998.

[Ras89] Rasmusen, Eric: *Games and Information : An Introduction to Game Theory.* Oxford Blackwell, 1989.

[RFH+01] Ratnasamy, Sylvia, Paul Francis, Mark Handley, Richard Karp und Scott Shenker: *A Scalable Content Addressable Network.* In: *Proceedings of ACM SIGCOMM 2001*, Seiten 161–172, 2001.

[RT99] Royer, Elizabeth und Chai-Keong Toh: *A Review of Current Routing Protocols for Ad-Hoc Mobile Wireless Networks.* IEEE Personal Communications, April 1999.

[Sab03] Sabater, Jordi: *Trust and Reputation for Agent Societies.* Doktorarbeit, Institute for Artifical Intelligence, Bellatera, Spanien, 2003.

[SAJ02] Seigneur, Jean-Marc, Joerg Abendroth und Christian Damsgaard Jensen: *Bank Accounting and Ubiquitous Brokering of Trustos.* In: *Proceedings of the 7th CaberNet Radicals Workshop*, Bologna, Italien, 2002.

[SAS95] Stanley, E. Ann, Dan Ashlock und Mark D. Smucker: *Iterated Prisoner's Dilemma with Choice and Refusal of Partners: Evolutionary Results.* In: *Advances in Artificial Life: Third European Conference on Artificial Life.* Lecture Notes in Artificial Intelligence, Springer-Verlag, 1995.

[SBHJ03] Salem, Naouel Ben, Levente Buttyan, JeanPierre Hubaux und Markus Jakobsson: *A Charging and Rewarding Scheme for Packet Forwarding in Multihop Cellular Networks.* In: *Proc. of the Fourth ACM Intl. Symposium on Mobile Ad Hoc Networking and Computing (MobiHoc'03)*, Annaplois, MD, USA, 2003.

[Sel78] Selten, Reinhard: *The Chain Store Paradox.* Theory and Decision, 9:127–158, 1978.

[Spe74] Spence, A. Michael: *Market Signaling: Informational Transfer in Hiring and Related Screening Processes.* Cambridge, Mass.: Harvard University Press, 1974.

[SS02a] Samuelson, Pamela und Suzanne Scothmer: *The Law and Economics of Reverse Engineering.* Yale Law Journal, Seiten 1630–1649, 2002.

[SS02b] Sen, Sandip und Neelima Sajja: *Robustness of Reputation-Based Trust: Boolean Case.* In: *Proceedings of the First International Joint Conference on Autonomous Agents and Multiagent Systems (AAMAS'02)*, Seiten 288–293, Bologna, Italien, 2002.

[SSA94] Smucker, Mark D., E. Ann Stanley und Dan Ashlock: *Analyzing Social Network Structures in the IteratedPrisoner's Dilemma with Choice and Refusal.* Technischer Bericht CS-TR-94-1259, University of Wisconsin-Madison, Department of Computer Sciences, 1994.

[ST97] Shoham, Yoav und Katsumi Tanaka: *A Dynamic Theory of Incentives in Multi Agent Systems.* In: *Proceedings of Fifteenth International Joint Conference on Arti Cial Intelligence (IJCAI) '97, Volume*, Seiten 626–631, 1997.

[SY01] Shionoya, Yoichi und Kiichiro Yagi: *Competition, Trust, and Cooperation: A Comparative Study.* Springer, 2001.

[Syv98] Syverson, Paul: *Weakly Secret Bit Commitment: Applications to Lotteries and Fair Exchange.* In: *Proceedings of the IEEE Computer Security Foundations Workshop*, Seiten 1–13, 1998.

[TCP00] TCPA: *Trusted Computing Platform Alliance Design Philosophies and Concepts*, 2000. http://www.trustedcomputing.org.

[Tuc50] Tucker, Albert: *A Two-Person Dilemma.* In: *Stanford University Press*, 1950.

[Tuo95] Tuomela, Raimo: *The Importance of Us: A Philosophical Study of Basic Social Norms.* Stanford University Press, Stanford, California, 1995.

[UMT03] UMTS, 2003. http://www.umts-forum.org.

[Usu02] Usunier, Jean-Claude: *An Open Electronic Bargaining System.* Technischer Bericht IUMI 0201, HEC Lausanne, 2002.

[Vas02] Vassileva, Julita: *Motivating Participation in Peer to Peer Communities.* In: *Proceedings of the Third Intl. Workshop on Engineering Societies in the Agents World (ESAW'02)*, Madrid, Spanien, 2002.

[Vid03] Vidal, Jose M.: *An Incentive-Compatible Distributed Recommendation Model.* In: *Proceedings of the Sixth International Workshop on Trust, Privacy, Deception, and Fraud in Agent Societies*, Seiten 84–91, 2003.

[Vid05] Vidal, José M.: *A Protocol for a Distributed Recommender System.* In: Falcone, Rino, Suzanne Barber, Jordi Sabater und Munindar Singh (Herausgeber): *Trusting Agents for Trusting Electronic Societies.* Springer, 2005.

[W3C04] W3C: *Web Service Architecture.* http://www.w3.org/TR/ws-arch, Februar 2004.

[WA95] Wu, Jianzhong und Robert Axelrod: *How to Cope with Noise in the Iterated Prisoner's Dilemma.* Journal of Conflict Resolution, 39(1):183–189, 1995.

[WJI04] Whitby, Andrew, Audung Josang und Jadwiga Indulska: *Filtering Out Unfair Ratings in Bayesian Reputation Systems.* In: *AAMAS'04-Workshop on Trust in Agent Societies*, New York, 2004.

[Woo02] Wooldridge, M.: *An Introduction to Multiagent Systems.* Wiley, 2002.

[ZEE03] Zyren, Jim, Eddie Enders und Ted Edmondson: *802.11g Starts Answering WLAN Range Questions*, 2003. http://www.commsdesign.com/story/OEG20030114S0008.

[ZJPV02] Zadeh, Ali N., Bijan Jabbari, Raymond Pickholtz und Branimir Vojcic: *Self-Organizing Packet Radio Ad Hoc Networks with Overlay (SOPRANO).* IEEE Communications Magazine, Seiten 149–157, Juni 2002.

[Zou05] Zou, Ying: *Entwicklung eines Hypothesen-Testers für simulierte Ad-hoc-Netze*, Mai 2005. Studienarbeit an der Fakultät für Informatik, Universität Karlsruhe.

Anhang

Anhang A

Aspekte der Implementierung und Evaluation

In diesem Anhang sind einige Details zur Implementierung und Evaluation des eigenen Ansatzes zu finden, die für das Verständnis der Arbeit zwar nicht notwendig aber hilfreich sind. Abschnitt A.1 befasst sich mit Aspekten der Implementierung, während die der Evaluation in Abschnitt A.2 abgehandelt werden.

A.1 Implementierung

A.1.1 Modellierung von Beweismitteln und ihrer Semantik im Regelsystem

In Abschnitt 7.5.2 sind wir darauf eingegangen, wie die Ablageverwaltung in [Dör04] implementiert ist. Ihr zugrunde liegt ein regelbasiertes System, dessen Einsatz die Spezifikation entsprechender Ableitungsregeln erfordert. Im Folgenden gehen wir auf diese Regeln ein. In ihnen ist implizit die Semantik von Beweismitteln und Wissen enthalten. Die Darstellung der Regeln bezieht auch die Eigenwechsel und Eigenwechsel-Invalidierungen aus den weiteren Konzepten des Abschnitts 12.1 mit ein. Dies wird durch den modularen Aufbau der Ableitungsregeln ermöglicht: Im *Kern* werden diejenigen Regeln definiert, die sowohl von transaktionalen Beweismitteln als auch den Eigenwechseln zugrunde liegen. Anschließend werden entsprechende Erweiterungen des Kerns vorgestellt.

In regelbasierten Systemen werden die Informationen, auf die die Ableitungen operieren, *Fakten* genannt. Diese beziehen sich bei unserem Systementwurf auf die Aussagen von Beweismitteln und das eigene Wissen. Zur Darstellung der Ableitungsregeln benötigen wir zunächst eine Definition der Struktur dieser Fakten. Sie stimmt mit der Tupeldarstellung aus Abschnitt 7.5.1 überein.

Der Kern. In der Tabelle A.1 wird die Struktur der Fakten definiert, die den Kern des Regelsystems bilden. Die Bezeichnungen sind wie folgt zu verstehen:

- *Arten von Beweismitteln:* PERMIT und ATTESTATION sind Verallgemeinerungen von Verträgen und Eigenwechseln einerseits und Quittungen und Eigenwechsel-Invalidierungen andererseits. Unter EVIDENCERID und ATTESTERID ist jeweils die Identität des Ausstellers des Beweismittels zu verstehen. CONTEXT bezieht sich auf die Beschreibung der versprochenen beziehungsweise ausgeführten Aktion. Eine negative Empfehlung wird DISRECOMMENDATION

Tabelle A.1: Fakten des Kerns des Regelsystems

- Permit:
  (EvidencerID, TransactionPeerID, EvidencerTransactionID, Context, InvalidationTime)
- Attestation:
  (AttesterID, AttesteeID, AttesteeTransactionID, Context, InvalidationTime)
- Disrecommendation:
  (EvidencerID, DisrecommendeeID, DisrecommendeeTransactionID)
- Cooperation:
  (CooperatorID, CooperatorTransactionID)
- Defection:
  (DefectorID, DefectorTransactionID)
- SelfRecommendationSource:
  (AttesterID, Context, InvalidationTime)
- DisRecommendationSource:
  (DisrecommendeeID, DisrecommendeeTransactionID, Context, InvalidationTime)
- TransferableDisRecommendationSource:
  (DisrecommenderID, DisrecommendeeID, DisrecommendeeTransactionID, Context, InvalidationTime)
- Inconsistency:
  (InconsistentEntityID)
- StatedReport:
  (StaterID, StateeID, Context, Report, InvalidationTime)
- MyEntityID:
  (EntityID)

bezeichnet. Sie wird von der Einheit der Identität EvidencerID über die Einheit der Identität DisrecommendeeID ausgestellt.

- *Arten von Wissen:* Cooperation und Defection beziehen sich auf die Wahrnehmung von Kooperation und Betrug. Außerdem gibt es einige Fakten, die die Möglichkeit zum Ausstellen von Empfehlungen anzeigen. Sie werden aus den entsprechenden zur Verfügung stehenden Beweismitteln abgeleitet. Bei diesen Fakten handelt es sich um SelfRecommendationSource für Selbstempfehlungen, DisRecommendationSource für negative Empfehlungen aufgrund eigener Erfahrungen, TransferableDisRecommendationSource für negative Empfehlungen, die bereits von Anderen erhalten wurden und weiter gegeben werden können, und Inconsistency für Typbeweise. Hinzu kommt das Fakt StatedReport, das der Glaubensbildung anzeigt, dass durch eine Quittung oder eine negative Empfehlung ein Bericht (etwa über den Ausgang einer Transaktion) erhalten worden ist. Report ist dabei wahr, wenn der Bericht dem Transak-

Tabelle A.2: Ableitungsregeln des Kerns des Regelsystems

- Permit(E_2,E_1,TID,C,T)
  $\wedge$ Disrecommendation(E_1,E_2,TID)
  $\wedge \neg$ MyEntityID(E_1)
  $\wedge \neg$ MyEntityID(E_2)
  $\longrightarrow$ StatedReport(E_1,E_2,C,FALSE,T)
- Attestation(E_1,E_2,TID,C,T_a)
  $\wedge \neg$ MyEntityID(E_1)
  $\wedge \neg$ MyEntityID(E_2)
  $\longrightarrow$ StatedReport(E_1,E_2,C,TRUE,T_a)
- Permit(E_1,E_2,TID,C,T)
  $\wedge$ Defection(E_1,TID)
  $\wedge$ MyEntityID(E_2)
  $\longrightarrow$ DisRecommendationSource(E_1,TID,C,T)
- Disrecommendation(E_1,E_2,TID)
  $\wedge$ Permit(E_2,E_1,TID,C,T)
  $\longrightarrow$ TransferableDisRecommendationSource(E_1,E_2,TID,C,T)
- Attestation(E_1,E_2,TID,C,T_a)
  $\wedge$ MyEntityID(E_2)
  $\longrightarrow$ SelfRecommendationSource(E_1,C,T_a)
- Disrecommendation(E_1,E_2,TID)
  $\wedge$ Attestation(E_1,E_2,TID,C,T_a)
  $\longrightarrow$ Inconsistency(E_1)

tionspartner kooperatives Verhalten bescheinigt.

- *Initiales Wissen:* Das Fakt MyEntityID ist initial dem Regelsystem hinzuzufügen. Es beinhaltet die eigene Identität. Dadurch ist es möglich zwischen Sachverhalten bezüglich sich selbst und Anderer zu unterscheiden.

Basierend auf dieser Definition der Fakten erhalten wir die Ableitungsregeln aus Tabelle A.2. Die nicht selbsterklärenden Teile der Regeln sind wie folgt zu verstehen:

- *Berichte:* StatedReport leitet eine Einheit nur dann ab, wenn sie am berichteten Vorgang unbeteiligt ist, wie dies zum Beispiel bei negativen Empfehlungen Anderer der Fall ist. Ansonsten bietet der Bericht nämlich keinen Informationsgehalt für die Glaubensbildung. Eine weitere Besonderheit entsprechend Abschnitt 7.3.1 besteht darin, dass eine negative Empfehlung nur dann als Bericht anerkannt wird, wenn sie von einem entsprechenden Vertrag begleitet wird.
- *Quellen von Empfehlungen:* Selbstempfehlungen und negative Empfehlungen stellt eine Einheit nur dann aus, wenn ihr gegenüber ein entsprechendes Permit (zum Beispiel Vertrag) oder eine Attestation (zum Beispiel Quittung) ausgestellt worden ist. Dies wird von den Verfahrensweisen gefordert. Negative Empfehlungen werden nur dann ausgestellt, wenn Wissen über den Betrug des Ausstellers des Permit vorliegt. Dadurch wird das Ausstellen inkonsistenter negativer Empfehlungen vermieden.

Tabelle A.3: Transaktionale Fakten des Regelsystems

- Contract:
  (EvidencerID, TransactionPeerID, EvidencerTransactionID, Context, InvalidationTime)
- Receipt:
  (EvidencerID, TransactionPeerID, PeerTransactionID, Context, InvalidationTime)
- NoAction:
  (DefectorID, DefectorTransactionID)

Tabelle A.4: Transaktionale Ableitungsregeln des Regelsystems

- Contract(e_1,e_2,tid,c,t)
  $\longrightarrow$ Permit(e_1,e_2,tid,c,t)
- NoAction(e,tid)
  $\longrightarrow$ Defection(e,tid)
- Receipt(e_1,e_2,tid,c,t)
  $\longrightarrow$ Attestation(e_1,e_2,tid,c,t)

Die Komponente der Glaubensbildung registriert sich somit auf die folgenden Ereignisse, um entsprechende Glaubensrevisionen anzusteuern:

- StatedReport (negativ): Konflikt (Abschnitt 7.4.2).
- Cooperation und Defection: Eigene Transaktionserfahrung (Abschnitt 6.3.3).
- Inconsistency: Typinformation und Rehabilitierung (Abschnitt 6.3.3 und Abschnitt 7.4.2).

Transaktionale Erweiterung. Der Kern des Regelsystems wird entsprechend der Tabelle A.3 für die Fakten und Tabelle A.4 für die Regeln erweitert. Diese Erweiterung führt eine eins-zu-eins Abbildung zwischen drei Paaren von Fakten durch:

- *Vertrag:* Permit und Contract
- *Quittung:* Attestation und Receipt
- *Fehlerhafte Aktionsausführung:* Defection und NoAction

Erweiterung um Eigenwechsel. Der Vorteil der Modularisierung des Regelsystems, die in der Definition des Kerns Ausdruck gefunden hat, wird deutlich, wenn wir das Regelsystem um Eigenwechsel erweitern. Wie zuvor bei der transaktionalen Erweiterung ist lediglich eine Abbildung durchzuführen. Diesmal lauten die Paare wie folgt:

Tabelle A.5: Fakten des Regelsystems bezüglich Eigenwechsel

- NOTE:
  (EVIDENCERID, CREDITORID, NOTEID, CONTEXT, INVALIDATIONTIME)
- NOTEINVALIDATION:
  (EVIDENCERID, DEBTORID, NOTEID, CONTEXT, INVALIDATIONTIME)
- NOTENOTHONORED:
  (DEBTORID, NOTEID)

Tabelle A.6: Ableitungsregeln des Regelsystems bezüglich Eigenwechsel

- NOTE(E_1,E_2,TID,C,T)
  $\longrightarrow$ PERMIT(E_1,E_2,NID,C,T)
- NOTENOTHONORED(E,TID)
  $\longrightarrow$ DEFECTION(E,NID)
- NOTEINVALIDATION(E_1,E_2,NID,C,T)
  $\longrightarrow$ ATTESTATION(E_1,E_2,NID,C,T)

- *Eigenwechsel:* PERMIT und NOTE
- *Entwertung:* ATTESTATION und NOTEINVALIDATION
- *Fehlendes Einlösen eines Eigenwechsel:* DEFECTION und NOTENOTHONORED

Tabelle A.5 und A.6 halten diese Paarbildung fest. Im Vergleich zu den transaktionalen Rahmenbedingungen, ist bei Eigenwechseln jedoch nicht von Transaktionspartner sondern vom Gläubiger (DEBTORID) und dem Schuldner (CREDITORID) die Rede. Der Schuldner ist derjenige, der dem Gläubiger seinen Eigenwechsel anvertraut hat. Damit transaktionale Beweismittel nicht in Konflikt mit Eigenwechseln geraten, müssen die Identifikatoren von Eigenwechseln (NOTEID) sich von denen für Transaktionen unterscheiden. Dies lässt sich durch ein entsprechendes Präoder Suffix erreichen.

A.2 Evaluation

A.2.1 Entwurfspunkt für die Evaluation

Laut Abschnitt 9.1.1 ist die Festlegung des Entwurfspunktes Voraussetzung für den Einsatz von Simulationswerkzeugen wie dem Interaktiven und Simulativen Kooperationsturniers. Bei der Beschreibung des Entwurfs in Teil II der Arbeit sind einige Entwurfsparameter offen gelassen worden, da ihre Quantifizierung von der jeweiligen Anwendungsdomäne abhängt. Diese Entwurfsparameter sind somit einzustellen. Das Kriterium hierfür ist ein grundsätzlich anderes als das für die Modellparameter der Rahmenbedingungen (aus Abschnitt 10.2.1): Während wir uns bei letzteren an der Realitätsnähe der Modellierung orientieren, sind die Entwurfsparameter so einzu-

stellen, dass die Güte der verteilten Vertrauensbildung optimiert wird. Bei der Festlegung der Entwurfsparameter muss daher der Standpunkt des Systementwerfers eingenommen werden.

Im Laufe von Vorversuchen im Interaktiven und Simulativen Kooperationsturnier hat sich eine Einstellung des Entwurfs als erfolgreich herauskristallisiert. Anhaltspunkt dabei waren die Eigenheiten der Umgebung des Campus-Szenarios. Diese Einstellung besprechen wir in diesem Abschnitt. Der sich ergebende Entwurfspunkt wurde bei der Durchführung der Evaluation in Kapitel 10 verwendet.

Die Entwurfsparameter werden in folgender Reihenfolge besprochen: Zunächst wenden wir uns der lokalen Vertrauensbildung aus Kapitel 6 zu. Anschließend werden die Entwurfsparameter der verteilten Vertrauensbildung eingestellt. Bei ihnen handelt es sich um die des Empfehlungsverhaltens (Kapitel 7) und der Bürgschaftsbeziehungen (Kapitel 8).

Lokale Vertrauensbildung. Es sind zwei Arten von Entwurfsparametern festzulegen. Sie betreffen einerseits das TIB-Modell und andererseits den Systemglauben.

Beim *TIB-Modell* sind die Wahrscheinlichkeiten für strategische Einhaltung p_n und unbeabsichtigtes Fehlverhalten p_u einzuschätzen. Bei der strategischen Einhaltung hat sich herausgestellt, dass strategische Einheiten keineswegs als immer betrügend angenommen werden können. Längerfristige Gegenstrategien sehen nämlich kooperatives Verhalten über weite Zeiträume vor. Diese Einsicht findet bei der Einschätzung von p_n Berücksichtigung, indem wir es als $(\frac{1}{2})^{0,2 \cdot v(\gamma)}$ definieren. Dabei gibt $v(\gamma)$ das Verhältnis des Transaktionswerts zum durchschnittlichen Transaktionswert an. Somit liegt die Einschätzung strategischer Einhaltung in der Gesamtevaluation des Abschnitts 10.2.1 je nach Transaktionskontext zwischen 80% und 93%. Diese Wahrscheinlichkeit muss deswegen so hoch gewählt werden, weil langfristig angelegte Gegenstrategien wie DISTRUSTDISTRIBUTOR eher selten Betrugsverhalten vorschreiben. Bei einer geringeren Einschätzung würden somit Einheiten, die solche Gegenstrategien verfolgen, als normativ erscheinen.

Die Wahrscheinlichkeit unbeabsichtigten Betrugsverhaltens lässt sich, wie in Abschnitt 6.3.2 vorgesehen, durch Messungen der Rahmenbedingungen ermitteln. Aufgrund dessen erhalten wir einen Wert von 15% für unbeabsichtigtes Betrugsverhalten in einer Transaktion. Auf den ersten Blick erscheint diese Einschätzung viel zu hoch, da in den Rahmenbedingungen lediglich von 5% unbeabsichtigter Fehler einer Aktionsausführung ausgegangen wird. Dieser Widerspruch löst sich auf, wenn wir berücksichtigen, dass in einer Transaktion außer der Ausführung zweier Aktionen zusätzlich Nachrichten mit Verträgen und Quittungen ausgetauscht werden. Der gemessene Wert ergibt sich bereits bei einer Fehlerrate von 1% pro Nachricht[1]. Wie hoch ist somit die Wahrscheinlichkeit p_u, dass einer der beiden Transaktionspartner sich unbeabsichtigterweise fehlverhält, einzuschätzen? Für diese muss $(\overline{p}_u)^2 = 85\%$ gelten. Daher wird im Entwurf p_u auf 7,8% eingestellt.

Auch der *Systemglaube* umfasst zwei Entwurfsparameter. Der initiale Systemglaube $p_X(N_0)$ bestimmt laut Abschnitt 6.3.3, wie eine Einheit X ihren Typglauben über die erste Einheit, die sie kennen lernt, setzt. Zudem gibt der Parameter α_0 an, welches Gewicht $p_X(N_0)$ in der fortwährenden Berechnung des Systemglaubens besitzt. Da wir als Systementwerfer a priori keine Information über den Anteil normativer Einheiten besitzen, hat sich die neutrale Einstellung $p_X(N_0) = 50\%$ als sinnvoll erwiesen. Bei der Festlegung des Gewichts ist zwischen einer schnellen Anpassung des Systemglaubens und der vorurteilsfreien Bewertung neuer Bekanntschaften

[1] Die genaue Berechnung hierzu ist wie folgt: In einer Transaktion müssen zwei Aktionen ausgeführt und (entsprechend des Sechs-Wege Protokolls) sechs Nachrichten übermittelt werden. Die Wahrscheinlichkeit, dass die Transaktion von Kommunikationsabbrüchen verschont bleibt, ist damit $(95\%)^2 \cdot (99\%)^6 \approx 85\%$.

abzuwägen. Die Einstellung $\alpha_0 = 5$ findet einen Kompromiss zwischen diesen Extremen.

Empfehlungsverhalten. Empfehlungen werden angefragt, wenn über potentielle Transaktionspartner zu entscheiden ist. Gemäß der Überlegungen aus Abschnitt 7.3.2 sind in diesem Moment eingehende Empfehlungen von besonderem Wert. Negative Empfehlungen über den potentiellen Transaktionspartner werden anhand des gemischten Algorithmus aus Abschnitt 7.3.2 eingeholt. Dabei werden bis zu maximal 5 Einheiten der unmittelbaren Umgebung über eine solche negative Empfehlung angefragt. Selbstempfehlungen werden vom potentiellen Transaktionspartner selbst angefragt, und zwar mit einer Wahrscheinlichkeit von 50%. Werden Selbstempfehlungen häufiger angefragt, steigt entsprechend der Aufwand, ohne dass die erhaltenen Selbstempfehlungen einen signifikanten Anteil bislang unbekannte Beweismittel enthalten würden. Bei einem zu seltenen Anfragen von Selbstempfehlungen werden hingegen wichtige Informationen (insbesondere bezüglich Bürgschaftsbeziehungen) verpasst. Anhand ähnlicher Überlegungen werden Typbeweise vom potentiellen Transaktionspartner zu einer Wahrscheinlichkeit von 20% angefragt. Diese Zahl ist geringer als bei Selbstempfehlungen, da sehr viel seltener neue Typbeweise als Bürgschaften und Quittungen anfallen.

Auch für die Beweismittel- und Wissensverwaltung, die das Empfehlungssystem unterstützt, ergibt sich ein Entwurfsparameter: Es ist die maximal zulässige Größe der Beweismittel-Basis. Laut Abschnitt 7.5.2 benötigt die Steuereinheit der Beweismittel- und Wissensverwaltung hierzu die Angabe, wie viele Beweismittel in der Beweismittel-Basis abgelegt werden können, bis es zur Verdrängung kommen soll. Für diesen Parameter hat sich gezeigt, dass eine Zahl von 50 vollkommen ausreichend ist. Bürgschaften werden aufgrund ihrer Wichtigkeit (und dem daraus resultierenden häufigen Zugriff) in der Regel nicht verdrängt.

Bürgschaftsbeziehungen. Die Entwurfsparameter der Bürgschaftsbeziehungen stehen in einem engen Zusammenhang. Dieser wurde bereits in Abschnitt 8.4.1 herausgestellt: Je stärker Bürgschaftsbeziehungen in der Glaubensbildung berücksichtigt werden, desto vorsichtiger muss eine Einheit beim Ausstellen einer Bürgschaft sein.

Bei der Einstellung des Entwurfs hat sich eine mittlere Berücksichtigung der Bürgschaftsbeziehungen als sinnvoll herausgestellt. Konkret bedeutet dies, dass die Entwurfsparameter zur Berechnung des sozialen Typglaubens durch $\lambda = \nu = 0,7$ angegeben werden. Ein geringerer Wert hätte die Bedeutung von Bürgschaften untergraben. Andererseits ist es bei höheren Werten zu einfach, die Folgen des eigenen schlechten Verhaltens auf die eigenen Bürgen abzuwälzen.

Außerdem sind die Parameter zum Eingehen von Bürgschaftsbeziehungen aus Abschnitt 8.4.1 zu setzen. Dabei ist zu berücksichtigen, dass die Bürgschaftsbeziehungen in der Glaubensbildung durch die Festlegung von λ und ν ein mittleres Gewicht haben. Demgemäß wird der minimale Typglaube, den eine Einheit zum Eingehen einer Bürgschaft verlangt, mit $p_\sigma = 80\%$ angegeben. Dadurch kommen einerseits normative Einheiten in vertretbarer Zeit zu Bürgschaftsbeziehungen. Andererseits ist für strategische Einheiten der Aufbau von Bürgschaftsbeziehungen kostspielig. Abschließend ist noch der Revisionsfaktor r_D festzulegen. Er wird wie im Beispiel des Abschnitts 8.3.2 auf $5,5$ gesetzt.