

# Vertrauen und Unternehmen

Werner Güth\* und Hartmut Kliemt\*\*

## Zusammenfassung

Bereits im Jahre 1980 wählte Horst Albach einen spieltheoretischen und experimentell-ökonomischen Zugang zum Vertrauensproblem in der Ökonomik. Dieser Weg wird erneut beschritten. Es wird aufgezeigt, dass neuere spieltheoretische und experimentelle Arbeiten nicht nur die explanativen Grenzen des ökonomischen Standardmodells opportunistisch rationalen Verhaltens verdeutlichen, sondern auch wesentliche Einsichten zur Rolle des Vertrauens in Organisationen bieten. Das wirft neues Licht auf die Leistungsfähigkeit, aber auch die Grenzen einer koordinativen Konzeption organisatorischer Zusammenarbeit, welches das Opportunismusproblem eher herunterspielt.

---

\* Max Planck Institut für Ökonomik, Jena

\*\* Frankfurt School of Finance and Management, HfB

## 1. Einleitung

Wenn ein Kind zum „Selters-Büchchen“ geht, um ein Eis zu erstehen, dann wird es sein Geld auf die Verkaufstheke legen und seinen Wunsch äußern. Es *verlässt* sich darauf, dass der Verkäufer nicht einfach die Münzen kassiert, ohne das sehnlichst erwartete Eis herauszurücken. Der Eisverkäufer hat als *Vertrauensnehmer* bereits die Vorleistung erhalten. Er könnte durchaus das Kind, das ihm gegenüber als *Vertrauensgeber* auftrat, um die Gegenleistung prellen.

Nicht nur Kinder, sondern alle Vertrauensnehmer müssen sich auf einen anderen verlassen. Sie sind insoweit verletzlich. Der Eisverkäufer wird dennoch das Vertrauen des Kindes kaum missbrauchen. Denn es würde sich erstens herumsprechen, wenn er derartiges wiederholt täte. Zweitens besitzen die meisten Eisverkäufer genügend Anstand, die von ihnen erwarteten Nachleistungen auch dann zu erbringen, wenn sie nicht Gefahr laufen, dass sich ein anderweitiges Verhalten herumsprechen könnte.

Was für Eisverkäufer gilt, lässt sich auf andere bilaterale, insbesondere stärker zeitversetzte, sowie auf multilaterale Beziehungen ausdehnen. Wenn beispielsweise drei Personen 1, 2, 3, miteinander Güter tauschen wollen und die Person 1 regelmäßig über das Gut A (benötigt von 3), 2 über Gut B (benötigt von 1) und 3 über Gut C (benötigt von 2) verfügt, dann könnten die drei auf die Idee verfallen, „Vertrauenspapiere“ herauszugeben. Diese („Geld“) Papiere „g“ helfen ihnen vor allem dann, den Tausch abzuwickeln, wenn sie nicht gleichzeitig über die Güter verfügen. Wo nicht direkt („Zug um Zug“) getauscht werden kann, nimmt jeder „erst einmal“ g entgegen, um später von einem anderen, der dafür seinerseits g entgegennimmt, etwas zu erhalten. Es handelt sich nicht um „do ut des“, wie in der einfachen bilateralen Austauschbeziehung,

sondern um „do ut det“ (vgl. dazu Weizsäcker, C. C. v. (1985), ergänzend Steiner, H. (1978)). Es könnte zu folgendem hypothetischem Ablauf kommen

1 hat g und erhält dafür B von 2

*2 nimmt g entgegen in der Erwartung, dafür C von 3 zu erhalten*

2 hat g und erhält dafür C von 3

*3 nimmt g entgegen in der Erwartung, dafür A von 1 zu erhalten*

3 hat g und erhält dafür A von 1

*1 nimmt g entgegen in der Erwartung, dafür B von 2 zu erhalten*

1 hat g und erhält dafür B von 2

*2 nimmt g entgegen in der Erwartung, dafür C von 3 zu erhalten*

2 hat g und erhält dafür C von 3

usw.

Dieses System funktioniert solange, wie jeder der drei erwartet, dass jeder andere weiterhin g entgegennimmt. Der *einzig*e Grund dafür, selbst g entgegenzunehmen, besteht darin, dass die anderen g entgegennehmen. So lange die anderen g akzeptieren, hat man einen guten Grund, g selbst zu akzeptieren. Wenn man den Verdacht hat, die anderen akzeptieren g nicht mehr, dann hat man hingegen guten Grund, selbst g nicht mehr zu akzeptieren. Wie im römischen Rundbogen (vgl. zu dieser Metapher auch, Binmore, K. (1992)) ein Stein den anderen stützt, so beruhen hier Erwartungen, dass g akzeptiert wird, auf Erwartungen (sowie Erwartungen von Erwartungen von...) und Vertrauen, dass man in g setzt, wenn man es entgegennimmt, stützt sich letztlich auf Vertrauen usw.

Äußere Interessenlagen und innere Neigungen spielen in allen wirtschaftlichen Transaktionen, in denen Leistungen und Gegenleistungen nicht unmittelbar miteinander verknüpft sind, eine ausschlaggebende Rolle. Vertrauen ist deshalb mit Recht ein zentrales Thema der heutigen Sozialtheorie. Bis zum Ende der 80-

er Jahre gab es dazu jedoch im Vergleich zur Bedeutung des Themas wenige Beiträge (mittlerweile gibt es insbesondere auch eine breite philosophische Literatur vgl. grundlegend und spieltheoretisch informiert Lahno, B. (2002)). Erst rückblickend erfahren die frühen Beiträge die ihnen gebührende Aufmerksamkeit. Ein Beispiel für den deutschen Sprachbereich bildet Horst Albachs mittlerweile viel zitierter Aufsatz „Vertrauen in der ökonomischen Theorie“ aus dem Jahre 1980 (vgl. Albach, H. (1980)). Aus Anlass eines Symposiums zum 75. Geburtstags von Heinz Sauermaun verfasst, nähert sich Albachs Diskussion dem Vertrauensthema auf der Basis spieltheoretischer und experimentell-ökonomischer Überlegungen. Die Wahl dieser Doppelperspektive mit Bezug auf das Thema des Vertrauens war bemerkenswert weitsichtig. Anlässlich des 75. Geburtstags von Horst Albach können wir dieser Weitsicht einen angemessenen Tribut zollen, indem wir uns erneut aus spieltheoretischer und experimentell-ökonomischer Perspektive mit Fragen des Vertrauens befassen und dabei den Interessen Horst Albachs entsprechend besonderes Augenmerk auf das Verhältnis von Vertrauen und Unternehmen richten. Ausgehend von Albachs Beitrag zu „Vertrauen in der ökonomischen Theorie“ (2.) werden wir einen spieltheoretischen Zugang zur Modellierung von Vertrauensbeziehungen skizzieren (3.), um im Anschluss daran einen exemplarischen Blick auf experimentell-ökonomische Ergebnisse zu Vertrauensspielen zu werfen (4.). Wir ziehen diese theoretischen und empirischen Ergebnisse schließlich heran, um zwischen einer koordinations- und einer konfliktorientierten Theorie des Unternehmens zu vermitteln (5.).

## **2. „Vertrauen in der ökonomischen Theorie“, drei Beobachtungen Horst Albachs**

In seinem Beitrag aus dem Jahre 1980 stellt Horst Albach fest, „auf gegenseitigem Vertrauen aufbauende Beziehungen zwischen Marktpartnern führen letztlich zu höheren Qualitäten, höheren Preisen und auch höheren

Gewinnen.“ (Albach, H. (1980), S. 5) Wenn auf Vertrauen aufbauende Beziehungen zu höheren Gewinnen für den Vertrauensnehmer führen, muss das allerdings *nicht* notwendig auf eine *wechselseitig* vorteilhafte Beziehung hindeuten. Es kann auch sein, dass ein Vertrauensnehmer das Vertrauen eines Vertrauensgebers ausbeutet. Die Chance zu einer opportunistischen Ausbeutung aufgebauten Vertrauens ist stets gegeben. Gleichwohl ist es nicht diese Form der Gewinnsteigerung, die Albach im Auge hat. Es geht ihm um Beziehungen, in denen Vertrauen auf eine entsprechende Gegenleistung und womöglich sogar auf „echte“ – im Gegensatz zu einer nur strategisch etwa mit Bezug auf Reputationsfolgen bestimmten – *Vertrauenswürdigkeit* trifft.

### **2.1. Von der Verlässlichkeit zum Vertrauen**

In vertraglichen Beziehungen, in denen die Vereinbarungen nicht vollständig spezifiziert sind und damit in allen praktischen Kontexten außerhalb der reinen ökonomischen Theorie, müssen entweder Misstrauen oder „Treu und Glauben“ herrschen. Auf das letztere konzentrieren wir uns im Weiteren. Ehrbare Kaufleute sehen sich auch und gerade an die stillschweigenden Übereinkünfte und das implizite Verständnis davon, was angemessen und dem Geschäftsbrauch entsprechend ist, gebunden. Insbesondere im Bereich von Dienstleistungen und dauerhaften Geschäftsbeziehungen, in denen es auf Verlässlichkeit ankommt und eine Bereitschaft, auch jenseits genau spezifizierter vertraglicher Pflichten zu kooperieren, ist man bereit, für diese höheren Qualitäten interpersonaler und inter-institutioneller Beziehungen höhere Preise zu zahlen. Wie so häufig führt gerade der Verzicht auf die Ausnutzung von kurzfristigen Vorteilen zu gesteigerten langfristigen Erträgen aller beteiligten Parteien.

Wenn die abdiskontierten langfristigen Erträge höher sind als die kurzfristigen Vorteile, dann hat der rationale Opportunist einen überzeugenden Grund, auf kurzfristige Versuchungen *nicht* einzugehen (vgl. etwa die grundlegenden

Beiträge in Klein, D. B. (1997)). Ginge er darauf ein, so wäre dies ein Rationalitätsmangel (vgl. zu solchen Rationalitätsmängeln, Ainslee, G. (2002)). Er könnte einer Versuchung nicht widerstehen und würde sich damit als letztlich nicht vertrauenswürdig zu erkennen geben. Der wahre Opportunist ist demgegenüber derjenige, der vollkommen vertrauenswürdig wirkt und an den entscheidenden Stellen, an denen man es nicht beobachten kann, genau in dem nicht feststellbaren bzw. für ihn maximal gewinnträchtigen Maß „mogelt“. Wer demgegenüber jede sich bietende Möglichkeit zu kurzfristigen Gewinnen wahrnimmt, der ist kein Opportunist, der ist nur dumm.

Auf einen langfristig klugen Akteur kann man sich in allen Kontexten verlassen, in denen das langfristige Eigeninteresse kurzfristig kooperatives Verhalten gebietet (alle auf das Folk-Theorem aufbauenden Konzeptionen rationaler Kooperation benutzen dies; vgl. dazu vor allem Aumann, R. J. (1981), Taylor, M. (1976), Taylor, M. (1987) und mit Bezug auf die betriebliche Situation Schauenberg, B. (1991)). Doch auch wenn sich rationale Akteure auf einen solchen Partner im Allgemeinen verließen, würden sie kaum in einem darüber hinausgehenden Sinne „vertrauen“. Das Vertrauen, das wir jemanden entgegen bringen, bezieht sich gerade auch auf jene Situationen, in denen er eigentlich gefahrlos und mit guten rationalen Gründen von jener langfristig klugen Strategie, auf die wir uns im großen und ganzen verlassen dürfen, abweichen könnte. Vertrauenswürdigkeit zielt auf etwas anderes ab. Der im engeren Sinne vertrauenswürdige Akteur ist selbst dann an innere Wertvorstellungen und Beschränkungen seines Opportunismus gebunden, wenn es sich unter Einbeziehung aller erwarteten (abdiskontierten) zukünftigen Folgen des einzelnen Aktes lohnen würde, eine Chance opportunistisch wahrzunehmen.

Worum es geht, lässt sich am Beispiel einer so genannten *goldenen Gelegenheit* illustrieren (vgl. zu goldenen Gelegenheiten Frank, R. (1992)). Bei einer solchen

Gelegenheit kann jemand das ihm entgegengebrachte Vertrauen ausnutzen, ohne dass eine realistische Gefahr der Entdeckung besteht. Wenn eine Kraftfahrzeugwerkstatt beispielsweise feststellt, dass der Stillstand des Motors durch den trivialen Ersatz eines Relais behoben werden könnte, dann bietet sich ihr womöglich die goldene Gelegenheit, umfangreichere Untersuchungen als die tatsächlich durchgeführten in Rechnung zu stellen. Der hintergangene Kunde der Kfz-Werkstatt hat praktisch keine Chance, die Sache aufzudecken. Verzichtet die Werkstatt dennoch darauf, dem Kunden eine größere Reparatur in Rechnung zu stellen, weil sie nach bestimmten Normen vorgeht, hat sie eine goldene Gelegenheit, deren Wahrnehmung dem langfristigen Interesse nicht widersprochen hätte, verstreichen lassen, ohne dass der Kunde überhaupt etwas davon bemerkte.

Beispiele wie das der Kfz-Reparatur zeigen, dass für ein vollkommen rationales Wesen die innere Gebundenheit gegenüber der Fähigkeit, jede sich bietende unter Einbeziehung aller Folgen vorteilhafte Chance wahrnehmen zu können, immer nachteilig ist. Wenn annahmegemäß alle zukünftigen Handlungsfolgen in die Erwartungswertbildung einbezogen wurden, ist es analytisch wahr, dass der Verzicht auf die erwartungswertmaximierende Handlung unmittelbar nachteilig sein muss. Die Diagnose der Nachteiligkeit ändert sich aber dann, wenn Handlungen aufgrund von festen Eigenschaften oder Dispositionen vollzogen werden. Wenn Vertrauenswürdigkeit und die inneren Bindungen, die dieser zugrunde liegen, als fixe (konstitutionelle) Dispositionen *erkennbar* sind, kann sich der Erwerb der *Disposition* zum Verzicht auf Vorteile lohnen. Zwar ist es nach wie vor nachteilig, die einzelne Goldene Gelegenheit verstreichen zu lassen. Aber, dass man eine Person ist, die generell goldene Gelegenheiten verstreichen lässt, kann womöglich vorteilhaft sein.

Wenn Individuen gebundene gegenüber ungebundenen Individuen als Transaktionspartner bevorzugen (vgl. dazu Baurmann, M. (1996) und in einem evolutionären Kontext ursprünglich Güth, W. and H. Kliemt (1993)), dann kann es einen Wert haben, gebunden zu sein. Voraussetzung dafür, dass sich *Tugend* in dieser Weise lohnen kann, ist, dass sich das Vorliegen der entsprechenden Disposition feststellen lässt. Das muss möglich sein aufgrund von Anzeichen, die jedenfalls teilweise unabhängig von dem Verhalten, auf das vertraut wird, beobachtbar sind. Eine solche Erkennbarkeit ist vermutlich vor allem in den dauerhaften inter-personalen Beziehungen innerhalb eines Unternehmens gegeben.

## **2.2. Vertrauen und Unternehmen**

In seiner Nobel-Lecture zur experimentellen Ökonomik und deren Aufstieg berichtet Vernon Smith ziemlich ausführlich über Experimente zum Vertrauen. Er motiviert die prominente Stellung, welche die Diskussion des Themas in seiner Vorlesung einnimmt, aufschlussreicher Weise wie folgt:

„I am reminded of a department head from Hewlett-Packard visiting our lab. I naively assumed that he would be most interested in demonstrations of some of our market experiments. Not so. He was more interested in the “trust” experiments. Why? He saw the HP management problem as one of getting teams to cooperate internally by building trust and trustworthiness, while being vigorous competitors externally. Could the trust games serve as a measurement and teaching tool for helping to solve this problem?” (Smith, V. L. (2003), 467, footnote 10)

Dem Hewlett Packard Managers geht es unter anderem um die alte Erkenntnis, dass vertrauensvolle Kooperation in einer Gruppe ein Mittel darstellt, in den Außenbeziehungen der Gruppe besonders wirkungsvoll zu konkurrieren. Das übersieht, dass Vertrauenswürdigkeit in externen Beziehungen ebenfalls ein Instrument der Konkurrenz sein kann. Vertrauenswürdigkeit in externen Geschäftsbeziehungen ist ein Politikparameter, den ein Unternehmen selber beeinflussen kann. Dazu kann es sich vor allem echte innere Bindungen durch



Satzungen oder Unternehmenswerte zulegen. Ob es bei Außenstehenden zum Vertrauen kommt, hängt aber nicht nur von der Existenz echter innerer Bindungen auf Seiten des Unternehmens als potentiellm Vertrauensnehmer in Interaktionen ab, sondern auch von hinreichender Zuverlässigkeit kognitiver Entdeckungs-Mechanismen auf Seiten potentieller Vertrauensgeber. Es verdient hervorgehoben zu werden, dass Horst Albach bereits in seinem Beitrag aus dem Jahre 1980 feststellte, „ich sehe einen wichtigen Beitrag der experimentellen Wirtschaftsforschung darin..., das sie aufzeigt, unter welchen Bedingungen es zulässig ist, Vertrauen als proxy für die mangelnde kognitive Kapazität einzusetzen. Je größer das Vertrauen ist, umso eher ist es dem rational Entscheidenden akzeptabel, die gegebenen kognitiven Grenzen zu respektieren.“ (Albach, H. (1980), S. 8)

Die eigenen kognitiven Beschränkungen machen unser Wissen in hohem Maße vertrauensabhängig. Wir reagieren besonders sensibel und ungehalten auf die Ausnutzung von Wissensasymmetrien durch Personen oder Institutionen, mit denen wir interagieren. Unternehmen müssen das im Umgang mit ihren Kunden und Lieferanten beherzigen, wenn sie auf Dauer erfolgreich sein wollen. Was das anbelangt darf die Strategie, Informationen zu teilen, als eine besonders wirksame Form der Signalisierung von Vertrauenswürdigkeit gelten. Als vertrauensbildende Maßnahme par excellence bietet es sich an, die Regeln jener internen Abläufe, nach denen Beschwerden von Kunden oder Lieferanten behandelt werden, transparent zu machen.

Ein Unternehmen kann nach der vorangehenden Überlegung innere Bindungen auf eine Weise signalisieren, die dem einzelnen Individuum nicht zugänglich ist. Insbesondere die internen Entscheidungsmechanismen des Homo oeconomicus befinden sich gleichsam in einer black box. Wenn es Bindungen des klassischen Homo oeconomicus gibt, dann sind diese immer externer Art. Er bindet sich

beispielsweise in Verträgen dadurch, dass er durch den Vertragsschluss für den Fall des Vertragsbruches zusätzliche Sanktionen negativer Art zu erwarten hat. Er ist nicht innerlich an den Vertrag gebunden. Seine Bindungen beruhen allein auf extrinsischen Anreizen. Im Umgang mit einem solchen Akteur geht es letztlich nur darum, dass man sich auf die Kombination aus Zielvorstellungen und deren rationaler Verfolgung *verlassen* kann.

Für die Wissenschaft der Ökonomik war es sicherlich einmal eine fruchtbare Arbeitshypothese, einen ausschließlich extrinsisch motivierten Akteur zu unterstellen. Dennoch zeigt uns insbesondere die experimentelle Spieltheorie, dass zumindest ein Modell, das sich allein an der einzelfallbezogenen Maximierung des Einkommens oder anderer materieller Gesichtspunkte orientiert, verfehlt sein muss. Spieler befolgen ganz offenkundig in ihren Interaktionen andere Maßstäbe, als es das normale Homo oeconomicus-Modell annehmen könnte. „Welche der Entscheidungen die Spieler treffen, hängt von dem Maß an Vertrauen ab, das die Spieler ineinander setzen. Und damit haben wir ihn ... wieder, den deus ex machina der ökonomischen Theorie, das Vertrauen“. (Albach, H. (1980), S. 6)

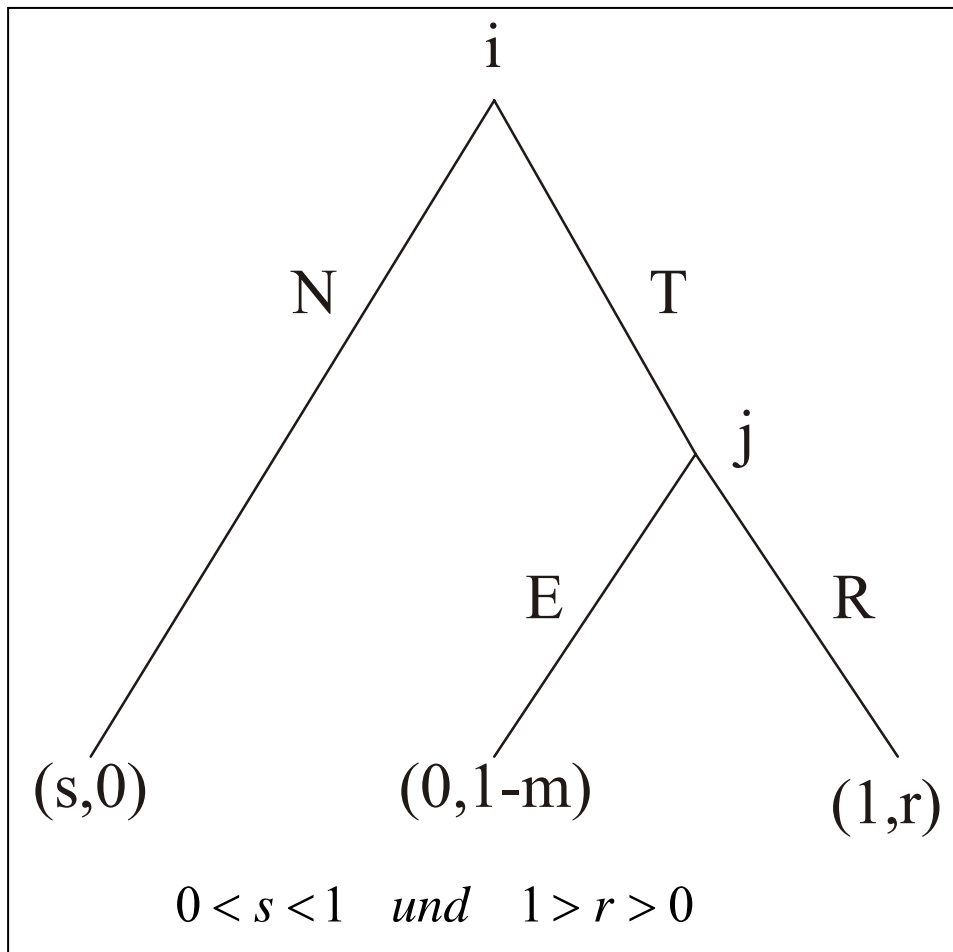
Mit einem „deus ex machina“ können wir uns auf die Dauer nicht zufrieden geben. Wir sind zwar noch weit von einer allgemeinen Theorie des Vertrauens insbesondere im Unternehmen entfernt, doch hat es seit Horst Albachs grundlegendem Aufsatz theoretische, empirische und auch normative Entwicklungen gegeben, die unser Verständnis des Vertrauens im unternehmerischen Handeln vertiefen können (vgl. zu der eng verknüpften Frage nach dem Verhältnis von intrinsischer und extrinsischer Motivation Frey, B. S. (1997), Le Grand, J. (2003)). Aus theoretischer (3.) und experimenteller (4.) Perspektive werden wir nun einige wenige dieser Entwicklungen exemplarisch

beleuchten, um daran grundlegende wenn auch vielleicht etwas spekulative Gedanken zur Rolle des Vertrauens im und für Unternehmen anzuschließen (5.).

### **3. Rationalwahl und Vertrauenswürdigkeit**

Ein bestimmtes Verhalten eines anderen Akteurs mit einer bestimmten Wahrscheinlichkeit vorauszusagen, ist etwas anderes, als dem anderen zu vertrauen. Dennoch muss man mit Verhaltensvoraussagen und deren Verlässlichkeit beginnen, wenn man sich ein einigermaßen zutreffendes Bild von Vertrauensbeziehungen machen will. Verlässlichkeit ist zwar keineswegs hinreichend für die Entstehung einer Vertrauensbeziehung. Verlässlichkeit ist aber eine Voraussetzung dafür, dass sich Vertrauen entwickeln kann.

Der nachfolgende einfache Graph gibt die Grundstruktur einer Interaktion wieder, in der Verlässlichkeit und Vertrauen eine Rolle spielen können:



Graphik 1

In diesem Archetypus einer „Vertrauensinteraktion“ muss ein Akteur  $i$  eine Entscheidung darüber treffen, ob er sich gegenüber dem Verhalten eines anderen Akteurs verletzlich machen will,  $T$ , oder nicht,  $N$ . Wählt er  $N$ , so ist das Ergebnis für ihn garantiert  $s$  mit  $0 < s < 1$  und für den anderen  $0$ . Wählt  $i$  hingegen  $T$ , so kommt  $j$  zum Zug. Entscheidet sich  $j$  für  $R$ , so erhält der Akteur  $i$  das von ihm am höchsten, mit  $1$ , bewertete Ergebnis, während der Akteur  $j$  für sich die Auszahlung  $r$  mit  $0 < r < 1$  realisiert. Wenn  $j$  sich hingegen nach  $T$  für  $E$  entscheidet, erhält Akteur  $i$  sein schlechtestes Resultat, das er mit  $0$  bewertet, während  $j$  mit  $1-m$  abschließt.

Die Repräsentation von Präferenzen durch Nutzenfunktionen lässt ohne Veränderung der Messaussage positiv affin lineare Transformationen zu. Diese erlauben eine Normierung der Zielfunktionen derart, dass das jeweils schlechteste Ergebnis den Wert „0“ und das jeweils beste Ergebnis den Wert „1“ in der Repräsentation der Präferenzen des jeweiligen Individuums erhält. Wir nehmen dabei an, dass außer dem Parameter  $m$  alle Parameter der Zielfunktionen der beiden Akteure  $i$  und  $j$  sowohl für objektive, materiale Auszahlungen (insbesondere monetärer Natur) als auch für subjektive Ergebnisbewertungen stehen. Der Parameter  $m$  hingegen hat kein objektives Gegenstück. Er repräsentiert alle jene nicht-monetären bzw. nicht-materiellen Gründe, die neben dem direkten objektiven Erfolg für den Vertrauensnehmer relevant sein und ihn entgegen seiner direkten materiellen Interessen zu einem im Sinne dieser Interessen nicht-opportunistischen Verhalten bewegen können.

Der rein „subjektive“ Charakter von  $m$  wird deutlich, wenn der Vertrauensnehmer  $E$  nach  $T$  wählen sollte. Dann ist seine materiale Auszahlung nämlich 1 während seine subjektive Bewertung der Situation durch  $1-m$  repräsentiert wird.

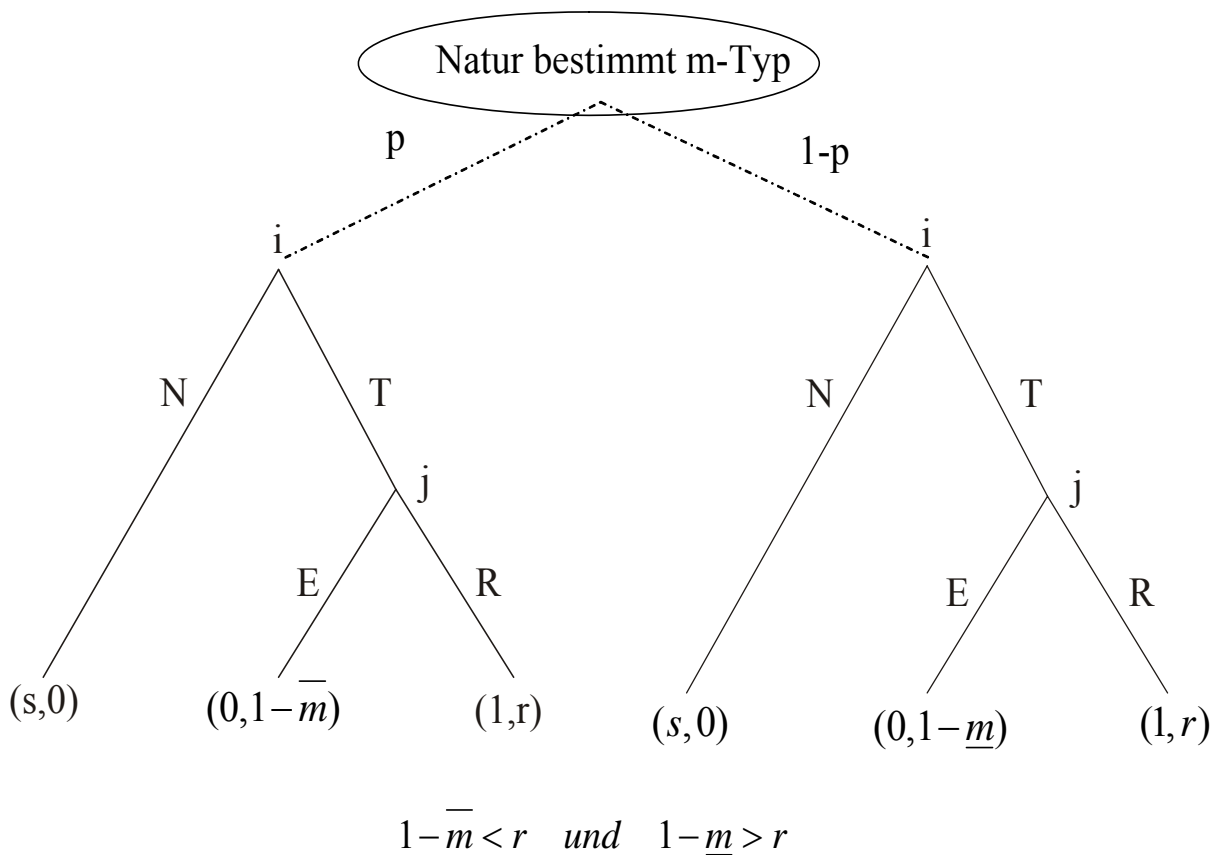
Wie man unmittelbar erkennt, wird in einer Situation wie der in Graphik 1 skizzierten bei vollständiger Information über Auszahlungen und Spielstruktur rationale Planung die Spieler zu teilspielperfekten Gleichgewichten führen mit:

$$r \geq 1-m \rightarrow (T, R)$$

$$r < 1-m \rightarrow (N, E)$$

Man bestimmt  $\bar{m} := 1-r$ . Alle Individuen mit  $m \geq 1-r = \bar{m}$  zeigen das gleiche Verhalten (der Einfachheit halber auch für  $r=1-m$  generell unterstellt). Sie werden daher als „ $\bar{m}$ -Typ“ bezeichnet. Alle Individuen mit  $m < 1-r = \bar{m}$  werden

zum „ $\underline{m}$ -Typ“ zusammengefasst. Wir nehmen zunächst an, dass ein Individuum  $i$  in der Rolle des Vertrauensgebers den Typ des Vertrauensnehmers  $j$  mit Sicherheit erkennen kann. Wenn der Anteil der Individuen, für die  $r \geq 1 - \underline{m}$  gilt, an der Gesamtpopulation  $p$  mit,  $1 \geq p \geq 0$ , beträgt und wenn es völlig zufällig ist, auf welche Individuen ein Akteur trifft, dann wird dieser, da er aufgrund der perfekten Typenkenntnis genau den vertrauenswürdigen Individuen vertrauen wird, in der Rolle des Vertrauensgebers mit Wahrscheinlichkeit  $p$  eine Auszahlung von 1 und mit Wahrscheinlichkeit  $(1-p)$  eine Auszahlung von  $s$  realisieren. Die nächste Graphik verdeutlicht die Situation bei perfekter Typen-Unterscheidbarkeit.



Graphik 2

Das bedeutet aber auch, dass es sich wegen  $r > 0$  für potentielle Vertrauensnehmer lohnt, vertrauenswürdig zu werden, falls sie ihren Typ selbst beeinflussen können. In einer Welt perfekter Typen-Unterscheidbarkeit würde

ein Homo oeconomicus, der das könnte, für sich eine innerer Gebundenheit entsprechende Zielfunktion zulegen (vgl. dazu Frank, R. (1987)). Er würde also aus einem *konstitutionellen* Eigeninteresse heraus (vgl. auch Vanberg, V. und J. M. Buchanan (1988)), seinen am materiellen Vorteil im Einzelfall orientierten Opportunismus zugunsten der Tugend der Verlässlichkeit beschränken.

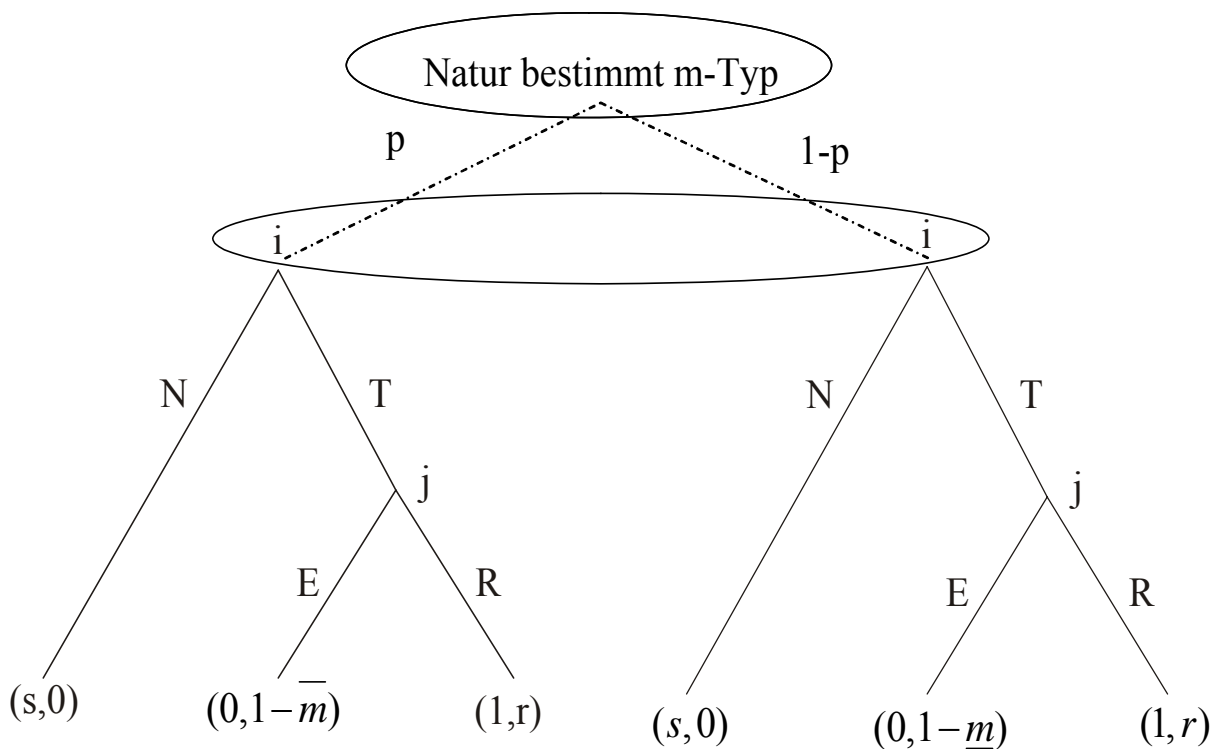
Wenn die Informationsbedingungen eine eindeutige und sichere Identifikation des Typs des Vertrauensnehmers nicht erlauben, dann ändern sich die grundlegenden Bedingungen für beide, Vertrauensnehmer und Vertrauensgeber. Für den Vertrauensnehmer schwächen sich die Anreize ab, die Tugend der Verlässlichkeit in sich zu kultivieren. Für einen Vertrauensgeber, der den  $m$ -Typ des Vertrauensnehmers nicht mehr sicher identifizieren kann, wird es zu einem riskanten Akt, T zu wählen.

Wenn der Typ des Vertrauensnehmers ausschließlich dessen private Information bleibt, wenn also vertrauenswürdige und nicht-vertrauenswürdige Typen für einen potentiellen Vertrauensgeber völlig ununterscheidbar sind, dann entsteht eine Situation, wie sie die nächste Graphik modelliert. Der Zufallszug der Natur drückt die durch die Populationszusammensetzung bestimmten Vermutungen des Vertrauensgebers bzw. dessen Unsicherheit über den Typ des anderen Spielers aus. Die Entscheidung für T bzw. N bestimmt sich nun nach dem Kriterium:

$$s > p1 + (1-p)0 \rightarrow N \text{ sollte gewählt werden}$$

$$p1 + (1-p)0 \geq s \rightarrow T \text{ sollte gewählt werden}$$

während der  $\bar{m}$ -Typ des Vertrauensnehmers typ-entsprechend R und der  $\underline{m}$ -Typ die Alternative E wählen wird.



$$1 - \bar{m} < r \quad \text{und} \quad 1 - \underline{m} > r$$

Graphik 3

Falls T gewählt wird, fahren ungebundene Individuen besser als die gebundenen. Wenn seine Gebundenheit wegen der vollkommen Ununterscheidbarkeit der Typen unerkennbar ist, kann der gebundene Akteur keine Vor-, sondern allenfalls Nachteile davon haben, gebunden zu sein. Wenn man das klassische Homo oeconomicus Modell zugrunde legt, ist es daher bei Typen-Ununterscheidbarkeit zwingend, dass ein Homo oeconomicus, der seinen Typ selbst strategisch wählen könnte, sich keinen gebundenen Typ aussuchen würde (zum vollen Spektrum von vollkommener Wahlmöglichkeit des Typs bis zur Unwählbarkeit, vgl. Berninghaus, S., W. Güth and H. Kliemt (2003)). Er würde sich dafür entscheiden, ein Homo oeconomicus zu sein, der jeden einzelnen Fall ohne jede Bindung an eine übergeordnete Regel nach dessen Meriten entscheiden kann.



Die Fälle perfekter Ununterscheidbarkeit ebenso wie die perfekter Unterscheidbarkeit sind Extreme, zwischen denen realistische Fälle einer partiellen Erkennbarkeit von Vertrauenswürdigkeit angesiedelt sind. In einer Welt, in der Verlässlichkeit mit einer gewissen Sicherheit diagnostiziert werden kann, sollte man sich unter bestimmten zusätzlichen Bedingungen wünschen, diese Eigenschaft zu besitzen (jedenfalls dann, wenn – wie hier unterstellt – der Besitz dieser Eigenschaft keine direkten Kosten verursacht). Ist der Besitz der generellen Disposition, nicht opportunistisch zu handeln, mit einer hinreichenden Sicherheit erkennbar, dann wird Gebundenheit indirekt vorteilhaft. Insoweit Dritte „Tugenden“ (i. e. innere Bindungen oder ein hinreichend großes  $m \geq \bar{m}$ ) zutreffend diagnostizieren und darauf strategisch reagieren können, wird es vorteilhaft, tugendhaft zu *sein* und nicht nur, so zu scheinen.

Das vorangehende Argument über die Vorteilhaftigkeit der Tugendhaftigkeit unter geeigneten Informationsbedingungen kann man unmittelbar in ein evolutionäres Modell überführen. Insbesondere die evolutionär stabilen Zusammensetzungen von Populationen konstitutionell verlässlicher (tugendhafter)  $\bar{m}$ -Typen und konstitutionell unverlässlicher  $\underline{m}$ -Typen lassen sich unter verschiedenen Informationsbedingungen und -kosten charakterisieren (vgl. dazu ausführlich z.B. Güth, W. and H. Kliemt (2000), Güth, W. and H. Kliemt (2006)). Derartige evolutionäre Modellierungen sagen etwas darüber, ob und unter welchen Bedingungen es theoretisch und „im Prinzip“ plausibel ist (vgl. zu Erklärungen im Prinzip auch Hayek, F. A. v. (1972), Stöckler, M. (1990), Stöckler, M. (2000)), dass „gebundene“ Nicht-Opportunisten in einer Welt, in der es auch „ungebundene“ Opportunisten gibt, überleben können. Da es aufgrund unserer Einschätzung der evolutionären Modellierungen sehr plausibel ist, dass echte Bindungen existieren, wollen wir im Gegensatz zu der Standardannahme der neo-klassischen Ökonomik für das weitere voraussetzen,

dass das Auftreten nicht-opportunistischen, gebundenen Verhaltens keine Anomalie im System unserer übrigen wissenschaftlichen Auffassungen darstellt, sondern im Rahmen der erwähnten und anderer Modellierungen (wie z.B. in Skyrms, B. (1996), Sober, E. and D. S. Wilson (1998)) eine zu erwartende Phänomen-Klasse darstellt. Das besagt aber noch nichts darüber, welche Phänomene man im Einzelnen zu erwarten hat. Wir wenden uns direkt den in Vertrauensinteraktionen unter kontrollierten Bedingungen zu beobachtenden Verhaltensphänomenen und damit der Frage zu, welches „typische Verhalten“ in experimentellen Studien zu Vertrauensinteraktionen auftritt.

#### **4. Experimente zu Vertrauenssituationen**

Die Experimente zu Vertrauenssituationen, die wir im folgenden ansprechen werden, lassen sich in drei Klassen unterteilen (vgl. zur experimentellen Ökonomik allgemein Kagel, J. H. and A. E. Roth (1995)):

##### **(i) Spiele mit binären Vertrauensentscheidungen**

Die binären Entscheidungen in Spielen dieses Typs laufen auf „Vertrauen zeigen oder nicht“, „Vertrauen belohnen oder nicht“ hinaus. Es liegt eine Auszahlungsdominanz der Strategiekombination „Vertrauen zeigen und Vertrauen belohnen“ über „kein Vertrauen zeigen und Vertrauen nicht belohnen“ vor. Das gemeinsame Interesse gebietet es, „Vertrauen zeigen und Vertrauen belohnen“ anzustreben, die opportunistische Rationalität führt jedoch, wie in Graphik 1 illustriert, zu der auszahlungsdominierten teilspielperfekten Lösung, „kein Vertrauen zeigen und Vertrauen nicht belohnen“ (vgl. zu einigen Experimenten dieser Art z. B. Güth, W., P. Ockenfels and M. Wendel (1992), Snijders, C. (1996)).

## (ii) Investitionsspiele mit Vertrauensmöglichkeiten

In diesen Interaktionen können Vertrauensgeber beliebige Beträge aus einem bestimmten Investitionsvolumen „gestuft“ in die Schaffung eines Gutes investieren, über dessen endgültige Verteilung der Vertrauensnehmer als Allokations-Diktator bestimmt. Vertrauensnehmer können Vertrauen in gestufter Form belohnen. Je mehr Vertrauen vom Vertrauensgeber gezeigt wird, umso höher ist die Steigerung der Effizienz, weil die Investition mit einem Multiplikator (größer als 1) zu einer Erhöhung der zu verteilenden Ressource führt. Der Multiplikator wurde in typischen Studien recht groß gewählt. Zum Beispiel wurde eine Verdreifachung jeder vom Vertrauensgeber weitergegebenen Geldeinheit untersucht (vgl. z. B. die grundlegende Studie Berg, J., J. Dickhaut and K. McCabe (1995)). Es lohnt sich also, von der Auszahlungsseite her, wenn beide Akteure von ihren teilspielperfekten Strategien abweichen. Der Vertrauensgeber macht sich jedoch durch die Investition gegenüber dem Vertrauensnehmer umso verletzlicher, desto mehr er investiert und der Vertrauensgeber ist zugleich umso stärkeren Versuchungen ausgesetzt.

## (iii) Wechselseitige Vertrauensspiele

In diesen Spielen kann jede der beteiligten Parteien die gemeinsamen Auszahlungen erhöhen, indem sie privat investiert. Sie macht sich jedoch jeweils dadurch verletzlich für die Ausbeutungsakte des jeweils anderen (vgl. dazu z. B. Berninghaus, S., L. G. González and W. Güth (2004)).

Neben diesen drei paradigmatischen Formen bieten noch viele andere Experimente Aufschluss über Vertrauensprobleme. Was unternehmensinterne Abläufe anbelangt, ist hier insbesondere an Principal-Agent-Experimente zu denken. Diese sind mit Vertrauensproblemen eng verknüpft. Externe Vertrauensfragen, wie sie sich in Beziehungen zwischen Unternehmen stellen,

werden durch vielfältige Marktexperimente ebenfalls in offenkundiger Weise angesprochen.

Wenn darüber hinaus Studien zur experimentellen Ökonomik (aber auch zur Psychologie) fast immer auch einen Vertrauensaspekt aufzuweisen scheinen, so liegt das einfach daran, dass menschliches Verhalten in strategischen Interaktionen generell in der Spannung von opportunistisch rationaler Verfolgung individueller und norm- bzw. regelgebundener Verfolgung gemeinsamer Ziele steht. Die Realisierung der Vorteile von Arbeitsteilung, Austausch und Zusammenarbeit erfolgt immer vor dem Hintergrund von Ausbeutungsmöglichkeiten und des so genannten Trittbrettfahrens (vgl. dazu grundsätzlich und elementar, Kliemt, H. (1993), Hardin, R. (1982)). Wir werden uns im Weiteren auf die drei vorgenannten, direkt dem Vertrauensthema gewidmeten Arten von Experimenten beschränken.

#### (i) Spiele mit binären Vertrauensentscheidungen

In den Experimenten von Güth, W., P. Ockenfels and M. Wendel (1992) wurden recht extreme Auszahlungsparameter gewählt. Diese ermöglichten hohe Effizienzgewinne bei gleichzeitiger Verstärkung der Ausbeutungsanreize. Mit diesen extremen Anreizen wurde

Vertrauen niemals belohnt und

Vertrauen ganz selten gezeigt.

Diese Ergebnisse zeigen, dass von intrinsischen Motiven geleitete Handlungen wie die, Vertrauen zu zeigen, beziehungsweise Vertrauen zu belohnen, von extrinsischen Anreizen überwogen werden können. Opportunismus wird durch intrinsische Motivationen nicht „koste es, was es wolle“, kontrolliert. Verhalten, das dem eines Homo oeconomicus zumindest nahe kommt, lässt sich somit induzieren (vgl. dazu auch Schumpeter, J. A. (1959)). Andere Experimente mit

weniger extremen Anreizen (vgl. Snijders, C. (1996)) zeigen aber, dass innere Bindungen häufig ausreichen, ein Verhalten herbeizuführen, in dem Vertrauen belohnt und auch vertrauendes Verhalten gezeigt wird. Diese Verhaltensbeobachtungen stehen im Widerspruch zum ökonomischen Standardmodell.

Der Standardversuch, diesen Widerspruch zum ökonomischen Verhaltensmodell zu neutralisieren, läuft darauf hinaus, auf Spiele von potentiell unendlicher Dauer zurückzugreifen (vgl. für einen Überblick zur Theorie infiniter Spiele Aumann, R. J. (1981)). Modelle wiederholter binärer Vertrauensentscheidungen eignen sich besonders zur Illustration dieses Vorgehens. Ein entsprechendes wiederholtes Vertrauensspiel unter den Partnern  $i$  und  $j$  der Graphik 1 (partner design) ergibt sich, wenn man sich beispielsweise das Spiel der Grafik 1 beliebig häufig so aufgeschrieben denkt, dass ein schleifenloser Graph mit einem definitiven Anfangs-, doch ohne bestimmten Endknoten entsteht. Der Spieler  $i$  beginnt und kommt nach  $T$  jeweils genau dann erneut zum Zuge, wenn der Spieler  $j$  die Alternative  $R$  in der vorangehenden Spielrunde wählte, während das Spiel endet, sobald  $j$  zum ersten Mal die Alternative  $E$  oder  $i$  die Alternative  $N$  wählt. Solange  $(T, R)$  gespielt wird, kommt es immer zu einer nächsten Runde.

In beliebig wiederholten Spielen könnte der Schatten der Zukunft (vgl. zu diesem Begriff natürlich Axelrod, R. (1987)) beliebig lang werden und damit ohne intrinsische Motivation zu erfordern, in jeder Runde mit Bezug auf spätere Interaktionen zu vertrauenswürdigem Verhalten anhalten. Menschliche Akteure wissen als endliche Wesen jedoch, dass sie nur über endliche Zeiträume miteinander interagieren. Für vollkommen rationale Akteure ist es daher gemeinsames Wissen, dass immer eine kleinste obere Zeit-Schranke existiert. Danach ist die Interaktion mit Sicherheit beendet.

Daraus ergibt sich eine fundamentale Schwierigkeit für alle solche Modelle unbegrenzt rationalen Verhaltens: Jenseits der kleinsten oberen Schranke gibt es keine Zukunft. Reputationseffekte sind irrelevant, weil der Schatten der Zukunft abgeschnitten wurde. Es kommt von der Schranke aus zur so genannten Rückwärtsinduktion. In der „Schrankenrunde“ haben rationale Akteure keinen Grund mehr, mit Bezug auf die Zukunft in vertrauenswürdiger Weise zu handeln. Da der Runde wird kein Vertrauen mehr gezeigt. Im gemeinsamen Wissen davon wird auch in der vorangehenden Runde weder Vertrauen gezeigt noch belohnt. Im gemeinsamen Wissen dieser Logik erreicht man schließlich die erste Runde und kommt zu dem Schluss, dass endliche Spiele dieser Art sogleich im Misstrauen enden sollten (vgl. Selten, R. (1978)).

Die Überlegungen zur Rückwärtsinduktion sind keineswegs nur technischer Natur, sondern treffen die Anwendbarkeit des ökonomischen Modells opportunistisch rationalen Verhaltens in seinem Kern. Wenn man nämlich beobachtbares Verhalten generell ohne die Annahme innerer Bindungen erklären will, muss man die Begrenztheit der Zukunft als *universelle* Anwendungsbedingung des Modells strikt zukunftsbezogenen Rationalverhaltens einbeziehen. Das gilt insbesondere auch für die Dauerhaftigkeit der Interaktion in einem Unternehmen. Obwohl typischerweise ausgedehnter als Marktinteraktionen ist auch diese Art der Interaktion immer begrenzt. Die Auffassung, dass man das für unternehmensinterne Abläufe charakteristische Ausmaß an Vertrauen und Vertrauenswürdigkeit gänzlich auf eine Internalisierung externer Effekte durch die stets gegebene Zukunft der Interaktion zurückführen könne, ist insoweit verfehlt. Für perfekt rationale endliche Wesen hat jede Zukunft eine kleinste obere Schranke, von der aus rückwärts geschlossen werden kann. Die Theorie der wiederholten Spiele hat

daher zwar einen wesentlichen Beitrag zum Verständnis rational kooperativen Verhaltens in Unternehmen zu bieten, reicht aber allein nicht aus.

Man hat versucht, dieser Konklusion zu entgehen, indem man vertrauensvolles Verhalten in endlich wiederholten Spielen auf Reputationseffekte zurückzuführen suchte (vgl. Kreps, D., P. Milgrom, J. Roberts and R. Wilson (1982)). Worum es geht, lässt sich wiederum am Beispiel der – nun allerdings endlich – wiederholten binären Vertrauensinteraktion erläutern. Beschränkt man sich z. B. auf eine fünfzigfache Wiederholung der binären Vertrauensinteraktion, so erhält man das bekannte „centipede game“ (Rosenthal, R. (1981); wollte man dem deutschen Ausdruck „Tausendfüßler“ gerecht werden, so müsste man das Spiel der Grafik 1 entsprechend 500-fach notieren, um 1000 Endknoten zu erzeugen). Hier müsste an sich die Rückwärtsinduktion greifen. Man unterstellt aber Typ-Unsicherheit. Man nimmt an, dass es einen bestimmten Anteil von Individuen gibt, die sich aufgrund ihrer Zielfunktion verlässlich in einer bestimmten „vertrauenswürdigen“ Weise verhalten. Wenn es solche Individuen gibt, dann reichen unter rationalen Erwartungen bei zunehmender Anzahl von Spielrunden zunehmend weniger solcher Individuen aus, um andere Individuen dazu zu bringen, sich anfänglich wie intrinsisch motivierte, nicht-opportunistische Individuen zu verhalten.

Das Argument hängt bezeichnenderweise davon ab, dass es wenigstens einige echt gebundene Individuen gibt. Es befindet sich insoweit gerade nicht in Übereinstimmung mit der Opportunismusannahme der Ökonomik, zu deren Verteidigung es doch ins Spiel gebracht wurde. Um zu sehen, ob abgesehen von diesem offenkundigen Einwand Reputationsmechanismen empirisch wenigstens näherungsweise in der unterstellten Form wirken, kann man wiederum Experimente durchführen.

Um die Präsenz echt gebundener Akteure in Experimenten sicherzustellen, kann man versuchen, einige Akteure durch Roboter zu ersetzen, die aufgrund des ihrem äußeren Verhalten zu Grunde liegenden Programms verlässlich in einer Weise agieren, die mit einem aufgrund von Vertrauenswürdigkeit gezeigten Verhalten übereinstimmt (und somit die Annahme unvollständiger Typeninformation rechtfertigen vgl. Anderhub, V., D. Engelmann and W. Güth (2002)). Die gleiche Übereinstimmung des äußeren Verhaltens von Interaktionspartnern mit einem aus intrinsischer Motivation erwachsenden vertrauenswürdigen Verhalten lässt sich auch durch spezielle monetäre Anreize erreichen. Man bietet dazu einer bestimmten Gruppe von Individuen finanzielle Belohnungen dafür, dass diese Individuen von der Ausbeutung anderer absehen. Sie erreichen durch eine Gegenleistung für gezeigtes Vertrauen eine höhere Auszahlung für sich selbst als durch die Ausbeutung (vgl. Güth, W., H. Kliemt, V. Levati und G. v. Wangenheim (2006 im Erscheinen)).

Verhaltensweisen, die dem entsprechen, was wir als Vertrauenswürdigkeit ansehen, mögen sich induzieren lassen. Es mögen auch einige der von den Anhängern einer universellen Anwendung der Opportunismusannahme vorausgesagten Effekte einer reputationsgelenkten Emulation der Vertrauenswürdigkeit eintreten. Die entsprechenden Verhaltensweisen erfordern jedoch komplexe Methoden der Bestimmung „gemischter“ Strategien beziehungsweise von Gleichgewichten in gemischten Strategien. Die experimentelle Evidenz kann die Hypothese, dass derartiges Mischungsverhalten in der Realität eine Rolle spielt, grundsätzlich nicht bestätigen. Insoweit lässt sich daher auch nicht behaupten, dass komplexe Theorien einer Bildung von Reputationsgleichgewichten durch spieltheoretische Experimente zu binären Vertrauensinteraktionen bestätigt werden.



Die Übereinstimmung zwischen den spieltheoretischen Reputationsgleichgewichten im centipede game und experimentell beobachteten Verhaltensweisen ist rein äußerlicher Natur: Wiederholung führt grundsätzlich zu mehr vertrauenden und zu mehr vertrauenswürdigen Verhaltensweisen in frühen Runden eines wiederholten Spiels, während zum Ende des Spiels so genannte Endspieeffekte opportunistisch rationalen Verhaltens auftreten (vgl. McKelvey, R. D. und T. R. Palfrey (1992) und zum Endspieeffekt in einem verwandten Kontext, Selten, R. und R. Stöcker (1983)). Doch Experimente zu elementaren Vertrauensinteraktionen der ersten Art – Vertrauensinteraktion vom Typ (i) – zeigen insgesamt, dass die äußere Übereinstimmung der Voraussagen von Modellen vollkommen rationalen Verhaltens mit den wahren verhaltenstheoretischen Ursachen der Beobachtungen nichts zu tun hat. Die Ursachen des beobachtbaren Verhaltens sind insbesondere nicht in den komplexen Überzeugungssystemen über Reputationsmechanismen, sondern zumindest auch in anderen, beschränkt rationalen Entscheidungsmaximen zu finden.

Die Theorie wiederholter Spiele bietet viele wertvolle Einsichten organisations-theoretischer Natur. Die entsprechenden (extrinsischen) Anreizeffekte bilden einen ausschlaggebenden Aspekt jeder Theorie des Unternehmens. Es wäre jedoch verfehlt, darüber die Rolle echter Vertrauenswürdigkeit und damit „gebunden rationalen Verhaltens“ auf der Grundlage intrinsisch motivierter Regelbefolgung im Unternehmen zu vernachlässigen. Das zeigt sich auch, wenn man einen näheren Blick auf Spiele richtet, die über binäre Entscheidungen hinaus Abstufungen von Vertrauen und Vertrauenswürdigkeit zulassen.

#### (ii) Investitionsspiele mit Vertrauensmöglichkeiten

Man stelle sich zwei Akteure, einen Vertrauensgeber  $i$  und einen Vertrauensnehmer  $j$  vor (vgl. zum folgenden Berg, J., J. Dickhaut and K.

McCabe (1995)). Der Vertrauensgeber und der Vertrauensnehmer sind beide im Besitz einer monetären Anfangsausstattung. Der Vertrauensgeber verfügt über  $X > 0$  und der Vertrauensnehmer über  $Y \geq 0$ . Begriffe wie „Vertrauen“ etc. wurden nicht benutzt, um die Regeln der Interaktion gegenüber den Akteuren zu beschreiben. Die Person, die in der Rolle eines Vertrauensgebers agieren sollte, konnte jeden beliebigen Teilbetrag  $x$ , mit  $0 \leq x \leq X$ , an die andere Person weitergeben und  $X - x$  für sich behalten. Wenn  $i$  einen Betrag  $x > 0$  vorsah, so sollte  $j$  den Betrag  $3x$  erhalten. Person  $j$  konnte dann wiederum jeden beliebigen Teilbetrag  $y \geq 0$  von  $3x$  an den ersten Akteur zurücksenden und  $Y + 3x - y$  für sich behalten.

Geht man beispielsweise einmal von den Beträgen von  $X = 10\text{€}$  und  $Y = 0\text{€}$  aus, dann würde sich  $X$  verdreifachen lassen, wenn der Akteur, der zunächst am Zug ist, den gesamten Betrag weitergeben würde. Der zweite Akteur könnte dann beispielsweise die Hälfte zurücksenden. In diesem Falle hätten der erste und der zweite Akteur beide  $15\text{€}$  erhalten. Das ist offenkundig eine Verbesserung für beide gegenüber einem Zustand, in dem nur der erste  $10\text{€}$  erhält und behält. Auf der anderen Seite könnte der zweite Akteur den gesamten Betrag, der ihm zugeht, für sich behalten. Er hätte dann  $30\text{€}$  und der erste Akteur, der es erst ermöglichte, dass der hohe Betrag verfügbar wurde, würde leer ausgehen.

Wie im einfachen binären Vertrauensspiel gibt es eine Vorleistung, die in Hoffnung auf Realisierung des gemeinsamen Vorteils erbracht wird, doch zugleich zu einer nur einseitig vorteilhaften Ausbeutung führen kann. Im Gegensatz zu den einfachen binären Vertrauensentscheidungen des vorangehenden Beispiels wird nun aber vom Vertrauensgeber nicht nur Vertrauen, sondern auch dessen Ausmaß zum Ausdruck gebracht. In den Experimenten zeigt sich, dass volles Vertrauen bzw. eine allein an Effizienz orientierte Weitergabe des gesamten zur Verfügung stehenden Betrages  $x = X$

generell eine höhere Belohnungsrate  $y/x$  induziert als die Weitergabe eines Betrages  $0 < x < X$ . Es macht psychologisch durchaus Sinn, dass jemand, der das Gefühl hat, dass ihm nicht voll vertraut wird, sich freier fühlt, selbst weniger vertrauenswürdig zu sein. Der Vertrauensnehmer kann in dem Experiment dementsprechend alle möglichen Arten von Abstufungen als Reaktion auf das ihm entgegengebrachte Vertrauen wählen. Er kann gar nichts zurücksenden ( $y=0$ ), ein durchaus nicht selten zu beobachtendes Verhalten, er kann dem Sender einfach dessen Einsatz  $x$  zurück überstellen, die Hälfte des bei ihm anlangenden Betrages (z.B.  $y=1.5x$ ) zurücksenden usw. Bezieht man mit ein, dass der Faktor 3 nur ein möglicher Multiplikator (größer als 1) ist und man auch die Zurücksendung mit einem Multiplikator versehen könnte, so ergeben sich weitere experimentelle Möglichkeiten, Vertrauensbeziehungen zu explorieren (vgl. mit betriebswirtschaftlichem Bezug etwa Ockenfels, A. (2003)).

### (iii) Wechselseitige Vertrauensspiele

In vielen gemeinsamen Unternehmungen müssen Beitragsentscheidungen getroffen werden, ohne dass es eine Sicherheit darüber gäbe, zu welchen Belohnungen diese Beiträge nach Beendigung der Unternehmung führen werden. Dies und die Trennung zwischen der Entscheidung über die Beitragserbringung und der Verteilung der Früchte der Zusammenarbeit kann generell dazu führen, dass die Beiträge entweder zu niedrig ausfallen oder sogar gänzlich entfallen. Das wird nur dann anders sein, wenn die Akteure Vertrauen zeigen und auch auf Vertrauenswürdigkeit stoßen. Wenn beispielsweise zwei Individuen  $i$  und  $j$  in firmenspezifisches Kapital investieren müssen, damit die volle Produktivität des Unternehmens gesichert werden kann, dann ergibt sich sogleich ein für wechselseitige Vertrauensspiele allgemein charakteristisches Spezialisierungsdilemma. Denn das firmenspezifische Kapital ist außerhalb des Unternehmens weniger wert als im Unternehmen. Wer also in firmenspezifisches Kapital schafft, der macht sich in dem Maße verletzlich, in dem

das Kapital an Wert außerhalb des speziellen Unternehmens verliert (zum theoretischen Hintergrund etwa Alchian, A. A. (1984), Milgrom, P. and J. Roberts (1993)). Wer sich z.B. so vollkommen spezialisiert, dass sein Kapital außerhalb des Unternehmens keinen Wert mehr besitzt, der riskiert es, auf einen weniger spezialisierten anderen zu treffen, dessen Kapital noch einen Wert außerhalb des Unternehmens hat. Der weniger Spezialisierte ist dann aufgrund seiner externen Optionen in einer stärkeren Verhandlungsposition (vgl. dazu theoretisch und experimentell Berninghaus, S., L. G. González and W. Güth (2004)).

Nahezu alle menschliche Arbeitsteilung und Spezialisierung ist, wiewohl im gemeinsamen Interesse, am Ende „antagonistische Kooperation“ (vgl. dazu Kliemt, H. (1986)). Neben der Hoffnung auf die Früchte der Zusammenarbeit enthält sie für alle Beteiligten das Risiko zum Opfer einseitiger Ausbeutung zu werden. Jedes gemeinsame Unternehmen bedarf der Vorleistung und des Vertrauens. Wo diese gegeben werden und auf Vertrauenswürdigkeit stoßen, entsteht der Eindruck, dass wir es unternehmerisch vornehmlich mit einer koordinativen Aufgabe zu tun haben. Eine rein koordinative Theorie der Unternehmung erscheint allerdings vor dem Hintergrund der vorangehenden theoretischen und experimentellen Überlegungen als ebenso extrem wie eine rein konflikttheoretische Konzeption.

## **5. Konflikt und Kooperation im Unternehmen**

Die Theorie der Vertrauensspiele ebenso wie die Experimente zu solchen Spielen werfen einiges Licht auf die Außenbeziehungen von Unternehmen zu ihren Kunden als auch die Binnenbeziehungen zwischen Mitarbeitern von Unternehmen. Um mit dem Letzteren zu beginnen, kann man sogleich an die vorangehenden Überlegungen zu wechselseitigen Abhängigkeiten und die darauf aufbauenden Vertrauensbeziehungen anknüpfen. Die weniger

spezialisierte Ressource kann ein strengeres Ultimatum glaubwürdig stellen. Diese Ressource ist schließlich außerhalb des Unternehmens und damit bei Wahrnehmung der Austrittoption relativ mehr wert. Wenn die Akteure aber rational die Druckmöglichkeiten in der Situation, in der es um die Verteilung der Früchte der gemeinsamen Anstrengungen geht, antizipieren, dann werden sie weniger spezifisch in die gemeinsame Anstrengung investieren und damit deren Früchte reduzieren.

Der wechselseitige Austausch von Pfändern oder „Geiseln“ mag in solchen Fällen hilfreich sein (vgl. Williamson, O. E. (1983), Raub, W. and G. Keren (1993)). Häufig wird es auch möglich sein, das Ausmaß oder den Grad der Ressourcenspezialisierung zu koordinieren. Eine Vermeidung von Ungleichgewichten in der wechselseitigen Abhängigkeit ist eine „vertrauensbildende“ und damit letztlich produktivitäts-steigernde Maßnahme. Bestimmte Ursachen für wechselseitige Verdächtigungen und *vorausseilendes Misstrauen* werden auf diese Weise in ihren Auswirkungen begrenzt. Dennoch wird man gerade in solchen Fällen nicht ohne ein gewisses Maß an Vertrauen auskommen und wird entstehendes Vertrauen die Effizienz der Kooperationen in hohem Maße fördern.

Reale Vertrauensbeziehungen in Unternehmen werden typischerweise nicht die einfache Form von binären Vertrauensentscheidungen annehmen und sie werden auch kaum die direkt wahrnehmbaren Effizienzvorteile des Investitionsspiels aufweisen. Insoweit werden sich experimentelle Ergebnisse nicht in direkter Form auf die Organisation von Interaktionen im Unternehmen übertragen lassen. Auch die wechselseitigen Ausbeutungsspiele, die im Hintergrund jeder Ressourcen-Spezifizierung stehen beziehungsweise durch diese latent erzeugt werden, bilden nur ein stilisiertes Abbild bestimmter Aspekte unternehmensinterner Interaktionen. Dennoch hat man es mit Elementar-Formen menschlicher

Interaktion zu tun, die exemplarisch verdeutlichen und analytischer Betrachtung zugänglich machen, was wir alle in unserem Alltagsleben erfahren: Jede Chance der Zusammenarbeit in gemeinsamen Projekten hat eine Gegenseite in den Risiken der Beitragsleistung.

Der Markt kann die Rolle persönlichen Vertrauens durch eine immer breitere Rolle von Reputationseffekten und/oder durch immer weitgehender spezifizierte Verträge reduzieren. Er kann damit eine Ausdehnung der Spezialisierung und Arbeitsteilung in immer anonymen Austauschbeziehungen ermöglichen. Das bedeutet jedoch nicht, dass er nicht Nischen dafür böte, in firmeninternen Abläufen den entgegengesetzten Weg zu beschreiten. In erfolgreichen Unternehmen gelingt es, den Bereich für wechselseitig verlässliche und vertrauensvolle Zusammenarbeit auszudehnen. Gegenüber dem Markt erweitert das Unternehmen den Bereich, in dem man ohne genaue Spezifikation von Leistung und Gegenleistung (koordinativ) zusammenwirken kann.

Wegen der omni-präsenten Vertrauens- und wechselseitigen Abhängigkeitsprobleme sind den nicht-marktlichen auf intrinsischen Motivationen aufbauenden Formen der Zusammenarbeit zwar Grenzen gesetzt. Als ganz wesentlicher Aspekt tritt jedoch „helfend“ hinzu, dass wir alle in dauerhaften Kleingruppeninteraktionen dazu neigen, unser stammesgeschichtliches Verhaltenserbe abzurufen. Das „Flüstern in uns“ führt uns dazu, dass wir in dauerhaften Kleingruppenbeziehungen eher aufgrund personaler und emotionaler Beziehungen als aufgrund nüchtern strategischer Betrachtungen handeln. Gut geführte Unternehmen sind in der Lage, unser tribales Erbe zu aktivieren und die menschliche Fähigkeit zu normgebundenem, intrinsisch motiviertem Verhalten „produktiv“ abzurufen. Zugleich sind sie allerdings darauf angewiesen, dass in anderen Hinsichten als Teil der Unternehmenskultur gerade die nüchterne Kalkulation von Vor- und Nachteilen vorherrscht.

Das arabische Sprichwort „ich gegen meinen Bruder, mein Bruder und ich gegen unsere Cousins, mein Bruder, unsere Cousins und ich gegen die anderen im Dorf, mein Bruder, unsere Cousins, die anderen im Dorf und ich, gegen die aus dem Nachbardorf usw.“ drückt die menschliche Elementarerfahrung, dass Kooperation ein Mittel der Konkurrenz sein kann, sehr gut aus. Erfolgreiche Unternehmen haben es vermocht, dieses Mittel besser zu nutzen als andere. Das spricht generell für eine Betonung koordinativer Elemente der Unternehmensführung: Wer aus Angst vor Kontrollverlust für die Vorteile des Vertrauens im unternehmerischen Leben blind wird, der wird davon große Nachteile haben. Er wird aus Mangel an Vertrauen die wesentliche Ressource der Vertrauenswürdigkeit, auf deren Existenz Experimente und Alltagserfahrung verweisen, nicht nutzen können und deshalb in der Konkurrenz mit anderen zurückfallen.

## Literatur

- AINSLEE, G. (2002). "Break Down of the Will." Princeton: Princeton University Press.
- ALBACH, H. (1980). "Vertrauen in der Ökonomischen Theorie." *Zeitschrift für die gesamte Staatswissenschaft*, 136, 2-11.
- ALCHIAN, A. A. (1984). "Specificity, Specialization, and Coalitions." *Zeitschrift für die gesamte Staatswissenschaft*, 140, 34 ff.
- ANDERHUB, V., D. ENGELMANN, and W. GÜTH (2002). "An Experimental Study of the Repeated Trust Game with Incomplete Information." *Journal of Economic Behavior and Organization*, 48, 197-216.
- AUMANN, R. J. (1981). "Survey of Repeated Games," in *Essays in Game Theory and Mathematical Economics*, ed. by R.J. e. a. Aumann. Mannheim: Bibliographisches Institut BI, 11-42.
- AXELROD, R. (1987). "Die Evolution der Kooperation." München und Wien: Oldenbourg.
- BAURMANN, M. (1996). "Der Markt der Tugend." Tübingen: Mohr.
- BERG, J., J. DICKHAUT, and K. MCCABE (1995). "Trust, Reciprocity, and Social History." *Games and Economic Behavior*, 10, 122-142.
- BERNINGHAUS, S., L. G. GONZÁLEZ, and W. GÜTH (2004). "Firm Specific Investments Based on Trust and Hiring Competition: A Theoretical and Experimental Study of Firm Loyalty." in *Schriftenreihe Des Wirtschaftswissenschaftlichen Seminars Ottobeuren*, ed. by W. Franz, Ramser, H.J., Stadler, M. Tübingen: Mohr Siebeck, 89-114.
- BERNINGHAUS, S., W. GÜTH, and H. KLIEMT (2003). "From Teleology to Evolution. Bridging the Gap between Rationality and Adaptation in Social Explanation." *Journal of Evolutionary Economics*, 13, 385-410.
- BINMORE, K. (1992). "Fun and Games - a Text on Game Theory." Lexington: D.C. Heath and Company.
- FRANK, R. (1987). "If Homo Economicus Could Choose His Own Utility Function, Would He Want One with a Conscience?" *The American Economic Review*, 77/4, 593-604.
- (1992). "Die Strategische Rolle der Emotionen." Oldenbourg: Oldenbourg Verlag.
- FREY, B. S. (1997). "Not Just for the Money. An Economic Theory of Personal Motivation." Cheltenham: Edward Elgar.
- GÜTH, W., and H. KLIEMT (1993). "Menschliche Kooperation basierend auf Vorleistungen und Vertrauen." in *Jahrbuch Für Politische Ökonomie*, ed. by P. e. a. Herder-Dorneich. Tübingen: Mohr.
- (2000). "Evolutionarily Stable Co-Operative Commitments." *Theory and Decision*, 49, 197-221.
- (2006). "Evolutionäre Spieltheorie in der Ökonomik." in *Evolution in Wirtschaft und Gesellschaft*, ed. by P. Weise. Marburg: Metropolis, 61-169.
- GÜTH, W., H. KLIEMT, V. LEVATI, and G. V. WANGENHEIM (forthcoming). "On the Co-Evolution of Retribution and Trustworthiness: An (Indirect) Evolutionary and Experimental Analysis." *Journal of Institutional and Theoretical Economics Zeitschrift für die gesamte Staatswissenschaft*, 2006.
- GÜTH, W., P. OCKENFELS, and M. WENDEL (1992). "Cooperation based on trust - An Experimental Investigation." *Journal of Economic Psychology*, 18, 15-43.
- HARDIN, R. (1982). "Exchange Theory on Strategic Basis." *Social Science Information*, 2, 251 ff.
- HAYEK, F. A. v. (1972). "Die Theorie komplexer Phänomene." Tübingen: Mohr.
- KAGEL, J. H., and A. E. ROTH (1995). "The Handbook of Experimental Economics." Princeton: Princeton University Press.
- KLEIN, D. B. (1997). "Reputation." Ann Arbor: The University of Michigan Press.
- KLIEMT, H. (1986). "Antagonistische Kooperation." Freiburg und München: Karl Alber.



- (1993). "Das Ricardosche Vergesellschaftungsgesetz und das Spezialisierungsdilemma." *WiSt*, 22., 335-341.
- KREPS, D., P. MILGROM, J. ROBERTS, and R. WILSON (1982). "Rational Cooperation in the Finitely-Repeated Prisoners' Dilemma." *Journal of Economic Theory*, 27, 245-252.
- LAHNO, B. (2002). "Der Begriff des Vertrauens." Paderborn: Mentis.
- LE GRAND, J. (2003). "Motivation, Agency, and Public Policy." New York et al.: Oxford University Press.
- MCKELVEY, R. D., and T. R. PALFREY (1992). "An Experimental Study of the Centipede Game." *Econometrica*, 60, 803-836.
- MILGROM, P., and J. ROBERTS (1993). "Economics, Organization and Management." Englewood Cliffs: Prentice Hall.
- OCKENFELS, A. (2003). "Reputationsmechanismen auf Internet-Marktplattformen." *Zeitschrift für Betriebswirtschaft*, 73, 295-315.
- RAUB, W., and G. KEREN (1993). "Hostages as a Commitment Device. A Game-Theoretic Model and an Empirical Test of Some Scenarios." *Journal of Economic Behavior and Organization*, 21, 43-67.
- ROSENTHAL, R. (1981). "Games of Perfect Information, Predatory Pricing, and the Chain Store Paradox." *Journal of Economic Theory*, 25, 92-100.
- SCHAUENBERG, B. (1991). "Organisationsprobleme bei dauerhafter Kooperation." in *Betriebswirtschaftslehre Und Ökonomische Theorie*, ed. by D. Ordelheide, B. Rudolph, and E. Büsselmann. Stuttgart: C. E. Poeschel, 329-356.
- SCHUMPETER, J. A. (1959). "The Theory of Economic Development." Cambridge, MA: Harvard University Press.
- SELTEN, R. (1978). "The Chain Store Paradox." *Theory and Decision*, 9, 127-159.
- SELTEN, R., and R. STÖCKER (1983). "End Behavior in Finite Prisoner's Dilemma Supergames." *Journal of Economic Behavior and Organization*, 7, 47-70.
- SKYRMS, B. (1996). "Evolution of the Social Contract." Cambridge: Cambridge University Press.
- SMITH, V. L. (2003). "Constructivist and Ecological Rationality in Economics." *American Economic Review*, 465-508.
- SNIJDERS, C. (1996). "Trust and Commitments." Amsterdam: Amsterdam.
- SOBER, E., and D. S. WILSON (1998). "Unto Others. The Evolution and Psychology of Unselfish Behavior." Cambridge and London: Harvard University Press.
- STEINER, H. (1978). "Can a Social Contract Be Signed by an Invisible Hand?," in *Democracy, Consensus and Social Contract*, ed. by P. Birnbaum, J. Lively, and G. Parry. London: Sage, 295-316.
- STÖCKLER, M. (1990). "Emergenz. Bausteine für eine Begriffsexplikation." *Conceptus*, XXIV, 7-24.
- (2000). "42 Jahre danach. Ein neuer Blick auf Oppenheim, Putnam und die Einheit der Wissenschaften." in *Argument Und Analyse*, ed. by C. U. Moulines, and K.-G. Niebergall. Paderborn: Mentis, 55-66.
- TAYLOR, M. (1976). "Anarchy and Cooperation." London u. a.
- (1987). "The Possibility of Cooperation." Cambridge: Cambridge University Press.
- VANBERG, V., and J. M. BUCHANAN (1988). "Rational Choice and Moral Order." *Analyse und Kritik*, 10/2, 138 ff.
- WEIZSÄCKER, C. C. v. (1985). "Zeit und Geld. Tertium Non Datur." Bern.
- WILLIAMSON, O. E. (1983). "Credible Commitments: Using Hostages to Support Exchange." *The American Economic Review*, 73/4, 519 ff.