

Werner Güth, Hartmut Kliemt und Stefan Napel

Wie Du mir, so ich Dir! – Ökonomische Theorie und Experiment am Beispiel der Reziprozität

Abstract

Economists usually treat human behavior as being determined by the shadow of the future, while most other social scientists point to the shadow of the past. This paper considers experimental evidence relevant to the controversy and tries to reconcile both explanations of human behavior with each other by integrating them in a unified evolutionary framework. The possible emergence and survival of intrinsically motivated resentment against being treated „unfairly” is analyzed as a case in point. The results shed light on the (in-)stability of different combinations of plain opportunism and social or ethical motives behind human behavior.

1. Einleitung und Übersicht

Bereits Adam Smith sprach davon, dass der Mensch von der Wiege bis zur Bahre nach einer Verbesserung seiner Lage strebe. Methodologische Normen der neo-klassischen ökonomischen Theorie verlangen ganz entsprechend, alle Entscheidungen als Opportunitäten ergreifendes Streben zu erklären. John Hicks (1979, 43) identifiziert dieses Streben als Kern des ökonomischen Verhaltensmodells mit den Worten: „people would act *economically*; when an

opportunity of an advantage was presented to them they would take it.” Nach dieser Sicht kennt die Ökonomik nur den Schatten der Zukunft. Alles Verhalten ist rational mit Bezug auf die erwarteten Folgen des Verhaltens. Der Ökonom hat eine akzeptable Erklärung für ein beobachtetes Verhalten nur und erst dann gefunden, wenn er es auf eine im zukunftsbezogenen Sinne rationale Entscheidung zurückführen kann. Wenn nicht, ist nicht die Theorie fragwürdig, sondern der Ökonom ist unfähig oder hat noch nicht lange genug nach einer ökonomischen Erklärung gesucht.

Eine Entscheidung ist *opportunistisch rational*, wenn sie aus einer Menge von Alternativen eine auswählt, die im Lichte der „materiellen“ Kausalfolgen der Wahl der Einzelalternative mindestens so gut abschneidet wie jede andere Alternative aus der Menge.

Aus Gründen begrifflicher Einfachheit haben wir den Begriff opportunistisch rationalen Verhaltens so definiert, dass er sowohl den Aspekt des Zukunftsbezuges als auch den einer rein „materiellen“ oder „objektiven“ Bewertung der Kausalfolgen umfasst. Es ist durchaus möglich, dass rationale Entscheider Kausalfolgen ihrer Verhaltensweisen nach subjektiven Maßstäben abweichend von den „objektiven Werten“ bewerten. Sequentielle Rationalität des Entscheidens kann mit Bezug auf im landläufigen Sinne „materielle“ (z.B. pekuniäre) wie „immaterielle“ (z.B. ethische) Bewertungen wirksam werden. So, wie wir den Opportunismusbegriff verwenden, muss jedoch beides zusammenkommen, erstens die sequentiell rationale Orientierung an den Kausalfolgen des jeweiligen zur Entscheidung anstehenden Aktes und zweitens die Bewertung dieser Kausalfolgen nach den klassischen materiellen Motiven etwa pekuniärer Art, die einen als intersubjektiv vergleichbar unterstellten Vorteil messen.

Opportunistisches Verhalten entspricht letztlich dem Referenz-Verhalten des klassischen ökonomischen Verhaltensmodells. Es ist allerdings wichtig, genau zu sehen, dass nach dem Modell opportunistischen, zukunftsgerichteten Verhaltens die Kausalfolgen des *einzelnen* Aktes zählen. Ob ein Akt einer Klasse von Handlungen zugehört, deren allgemeine Durchführung wünschenswerte Folgen hätte, ist nicht relevant. Ob es schlechte Folgen *hätte* wenn ich einen Akt einer bestimmten Art immer wieder durchführen würde, ist

irrelevant, wenn die Durchführung des einzelnen Aktes für mich besser ist als seine Unterlassung. Ob es schlechte Folgen *hätte* wenn alle einen Akt einer bestimmten Art durchführten, kann ebenfalls vernachlässigt werden. Ein opportunistisches rationales Individuum rechnet jeder Handlung genau die von dieser Handlung ausgehenden erwarteten Kausalfolgen zu und bewertet jede Entscheidung je für sich nach den Kausalfolgen der Einzelentscheidung. Verhaltensdispositionen, Tugenden, normorientiertes Verhalten, das von vergangenen Akten und Erfahrungen, nicht von Zukunftserwartungen bestimmt wird, fügen sich jedenfalls nicht auf natürliche Weise in dieses Bild.¹

Der Aufstieg der Ökonomik zu einer Sozialwissenschaft mit allgemeinem, nicht nur auf im engeren Sinne ökonomische, insonderheit Markt-Interaktionen beschränktem Erklärungsanspruch verdankt sich wesentlich der konsequenten Anwendung des Modells opportunistischen rationalen Entscheidens auf alle Lebensbereiche. Aufgrund der A priori-Annahme durchgängig opportunistischer rationaler Verhaltenssteuerung hat die Ökonomik gegenüber den anderen Sozialwissenschaften, die jeweils ein Konglomerat recht disparater Theorien bilden, bislang den großen Vorzug eines scheinbar einheitlichen methodischen Ansatzes für sich. In allen Bereichen sozialer Interaktion soll der Homo oeconomicus agieren und damit das gleiche Modell opportunistischen rationalen Verhaltens zugrunde gelegt werden können.

Die methodologische Kritik an der entscheidungslogisch orientierten Ökonomik, sie laufe auf Modellplatonismus anstatt empirische Erklärung hinaus, ist wohlbekannt (vgl. Albert, 1967). Dieser methodologischen Kritik kann sich der

¹ Die tiefere philosophische Wurzel etwa für das Phänomen nicht glaubwürdiger Drohungen oder Versprechungen liegt hier (vgl. zu den verwandten Perfektheitsproblemen Selten, 1965 und 1975). Der eigentliche Spieler der Spieltheorie ist, nimmt man das Konzept des Rationalverhaltens und insbesondere des opportunistisch rationalen Verhaltens beim Wort, nicht der personale Spieler, sondern der Agent der Agentennormalform, der die elementare unabhängige Entscheidungsinstanz bildet (vgl. Kuhn, 1953, und ausführlicher zu einigen der eher „philosophischen Fragen“, die sich in diesem Zusammenhang stellen, Güth und Kliemt, 1995a und b). Damit ist das ökonomische Verhaltensmodell jedoch weit von dem entfernt, was die übrigen Sozialtheoretiker für empirisch relevant halten würden. Es steht außer Frage, dass das ökonomische Modell, solchen Konzepten, wie dem der personalen Identität, des Personalen Spielers oder Entscheiders nicht leicht gerecht werden kann bzw. der Ökonom alle diese Eigenschaften in die explizite Modellierung von Spielregeln (zu denen ja auch die Präferenzen gehören) aufnehmen muss. Wenn er aber zugleich die Spielregeln als gegeben hinnehmen möchte, dann fragt sich, was noch als Gegenstand der Theorie bleibt.

Ökonom unter dem Eindruck empirischer, insbesondere experimenteller Befunde kaum verschließen. Er muss von der reinen Lehre abgehen, wonach alles Handeln im Lichte der Kausalfolgen jeder Einzelhandlung als Ausfluss optimierenden zukunftsgerichteten Entscheidungsverhaltens zu erklären ist. Neben dem Schatten der Zukunft und der Fähigkeit, sich bietende Opportunitäten zu ergreifen, müssen auch Ökonomen in ihren Erklärungen dem Schatten der Vergangenheit und der Fähigkeit, sich an Regeln und Prinzipien zu binden, Rechnung tragen. Damit ist die methodologische Zentralfrage gestellt, wie man der Existenz nicht-opportunistischen Verhaltens Rechnung tragen kann, ohne dem von Ökonomen viel gescholtenen soziologischen Modell einer Verhaltenssteuerung nach internalisierten Normen und Werten so viele Konzessionen zu machen, dass von der klassischen Ökonomik nichts mehr übrig bleibt bzw. unter dem Dach der Ökonomik zwei vollkommen verschiedene Ansätze verfolgt werden.

Bei der Formulierung von Antworten auf diese methodologische Herausforderung ist sorgfältig darauf zu achten, dass nicht das „Kind mit dem Bade“ ausgeschüttet wird. Denn es mag zwar sein, dass opportunistisch rationales Verhalten kein *universell* geeignetes Erklärungsmodell darstellt. Zur gleichen Zeit ist es vollkommen unangemessen, die menschliche Fähigkeit zu derartigem Verhalten völlig außer Acht zu lassen. Wenn menschliches Verhalten sowohl von der Vergangenheit ohne intervenierende Entscheidung „direkt kausal“ bestimmt und damit gleichsam „von hinten geschoben“ als auch im Lichte seiner erwarteten Kausalfolgen in bewussten strategischen Entscheidungen gewählt und damit gleichsam „von vorn gezogen“ wird, dann ist beidem Rechnung zu tragen.²

² Wir können hier nicht weiter auf die beiden ganz unterschiedlichen Arten der Abweichung vom opportunistischen Verhaltensmodell eingehen. Nur soviel sei gesagt: Da Opportunismus durch die zwei Elemente der Orientierung an materiellen Bewertungen und der Orientierung an der Einzelentscheidung charakterisiert ist, bedeutet dessen Negation, dass das eine oder das andere nicht zutrifft. Das Verhalten kann also an nicht materiellen Motiven orientiert und deshalb nicht-opportunistisch sein oder eine Regelbindung, die über den Einzelfall hinaus geht, beinhalten und deshalb die Ergreifung sich bietender Opportunitäten unmöglich machen. Dabei ist zu beachten, dass die Bindung an eine Regel durchaus aus materiellen Motiven erfolgen kann und damit dieses klassische Element des Homo oeconomicus Modells vorliegt, jedoch das andere fehlt. Wenn beispielsweise ein Akteur einem anderen droht, um dadurch seine materiellen Ziele besser erreichen zu können, dann muss die Drohung glaubwürdig sein. Unter opportunistischen Individuen ist sie das unter der Annahme, dass der Opportunismus

Der von uns am konkreten Beispiel der Reziprozität skizzierte indirekte evolutionäre Ansatz bietet einen Rahmen, in dem der Schatten der Zukunft und der Vergangenheit, untergebracht werden können. Im Folgenden werden wir zunächst das Konzept der Reziprozität charakterisieren (Abschnitt 2.1) und seine Wirkungsweise am „Ultimatumspiel“ illustrieren (Abschnitt 2.2). Im nächsten Schritt wird der indirekte evolutionäre Ansatz auf Ultimatumspiele angewendet, wobei in einem ersten Teilschritt (Abschnitt 3.1) von Informationsproblemen abgesehen wird, die im zweiten Teilschritt (Abschnitt 3.2) ebenfalls „endogenisiert“ werden. Es wird dann überlegt, in welcher Weise man den indirektevolutionären Prozess so erweitern könnte, dass er „generalisierte“ Dispositionen zu reziprokem Verhalten in heterogenen Umgebungen zu erfassen vermag, in denen strukturell verschiedene Spieltypen vorkommen (Abschnitt 3.3). Abschließende Bemerkungen (Abschnitt 4) verdeutlichen nochmals den programmatischen Charakter der hier angestellten Überlegungen und versuchen, diese in den Gesamtzusammenhang methodologischer Überlegungen der Sozialtheorie einzuordnen.

2. Reziprozität

gemeinsames Wissen ist, jedoch unter Umständen nicht. Wenn man nun einem Individuum, das bei opportunistisch rationalem Verhalten nur unglaubliche Drohungen aussprechen kann, die Möglichkeit bietet, seine Drohung durch einen Mechanismus (in Analogie zum Mast, an den sich Odysseus zum Schutz vor seiner erwarteten Schwäche angesichts der Lockgesänge der Sirenen binden ließ) glaubwürdig zu machen, dann würde dieses Individuum diese Selbstbindungsmöglichkeit auch bei einer reinen Orientierung an materiellen Folgen gern wahrnehmen wollen. Die Fähigkeit zur Selbstbindung wird dann wahrgenommen aus einem materiellen Motiv. Der Akteur kann in seinen Wahlakten ausschließlich materiell motiviert und im klassischen Sinne „eigeninteressiert“ sein. Zur besseren Wahrung dieses Eigeninteresses weicht er jedoch vom opportunistischen Verhalten ab. Dieser Fall der Beschränkung des Opportunismus aufgrund einer „reinen Regelorientierung“ ist ganz offenkundig verschieden von einer Umorientierung durch nicht-materielle Motive, denn, das sei wiederholt, die Regelbindung erfolgt gerade aus dem Wunsch, die eigenen materiellen Ziele besser zu verwirklichen. Die vorangehende Bemerkung spricht natürlich erneut nur explizit die „Philosophie“ hinter Reinhard Seltens inhaltlichen – im Gegensatz zu den technischen – Überlegungen zur (Teilspiel-)Perfektheit aus; vgl. für eine ausführlichere, noch wortreichere philosophische Analyse Kliemt (1993) und für extrem umfangreiche aber auch äußerst aufschlussreiche soziologische Diskussion Baumann „Der Markt der Tugend“ (1996).

Das Beispiel der Reziprozität ist nicht nur von methodologischem, sondern von systematischem Interesse, da alle Elementarformen menschlicher Kooperation und Organisation in der einen oder anderen Weise von Reziprozität abhängen. „Wie Du mir, so ich Dir!“ ist eine sprichwörtliche Maxime menschlichen Verhaltens. Doch obschon sich diese Maxime im Folk-Theorem (im einfachsten Fall der bekannten Tit For Tat-Strategie) anscheinend als Ausfluss sequentiell rationaler Entscheidungsfindung rechtfertigen lässt, zeigen Experimente, dass sie motivational eher auf retributive Emotionen und damit den Schatten der Vergangenheit als auf zukünftige rationale Erwartungen oder den Schatten der Zukunft zurückgeht.

Reziprozität hat sich offenkundig in früheren realen Evolutionsprozessen (biologischer oder sozialer Art) bewährt. In heutigen komplexen rechtlich organisierten Gesellschaften erlaubt sie es, die Früchte der Zusammenarbeit und die Früchte der Kooperation auch jenseits der Grenzen von expliziten Verträgen zu ernten. Es ist allerdings eher rätselhaft, wie sich eine entsprechende Tendenz zu reziproken Präferenzen oder auch komplett nicht-opportunistischem reziprokem Verhalten durchsetzen konnte. Wieso gelingt es insbesondere rein materiell orientierten Opportunisten nicht, die Nicht-Opportunisten zu verdrängen bzw. welche Bedingungen müssen vorliegen, damit sich im Rahmen zukunftsgerichteten Verhaltens eine Präferenz für reziprokes Verhalten als vorteilhaft insbesondere gegenüber einer an den rein materiellen Konsequenzen eigenen Handelns orientierten Verhaltenssteuerung erweist? Warum setzen sich nicht jene durch, die gerade nicht aufgrund genereller Einstellungen handeln, sondern jede sich bietende Gelegenheit zur Wahrung ihres materiellen Wohles wahrnehmen? Wie kann es sein, dass etwa eine generelle „intrinsische Motivation“ zur Reziprozität, die gerade gegen die materiellen Interessen im Einzelfall gerichtet ist, sich durchsetzen bzw. unter Konkurrenzbedingungen halten kann?

2.1. Zum Begriff der Reziprozität

Der Gemeinplatz „Wie Du mir, so ich Dir!“ ist ein guter Wegweiser, wenn es darum geht, den Kern des Reziprozitätskonzeptes zu erfassen. Gleiches wird mit Gleichem vergolten. Beeinträchtigungen werden mit „Bestrafungen“ oder

„negativen Sanktionen“, Verbesserungen mit „Belohnungen“ bzw. mit „positiven Sanktionen“ beantwortet.

Die diesen Sanktionsneigungen zugrunde liegenden retributiven Emotionen (vgl. Mackie, 1985, und Strawson, 1962) sind so tief in uns verankert, dass wir sie nicht nur gegenüber Personen zu zeigen pflegen. Der notorische Fall jenes Perserkönigs, der nach dem Untergang seiner Flotte das Meer auspeitschen ließ, zeigt, wie weit wir uns von retributiven Emotionen leiten lassen können. Im Mittelalter wurden Pferde, die eine menschliche Person zu schwerem Schaden gebracht hatten, deshalb förmlich hingerichtet. Auch in der heutigen Zeit gibt es durchaus noch ähnliche Verhaltensweisen. Doch sollten wir wohl von diesen extremen, eher pathologischen Verhaltensformen absehen und minimale Standards angemessenen Verhaltens zu Grunde legen, wenn wir Studien typischen menschlichen Verhaltens durchführen wollen.

Reziprozität im engeren Sinne wird in inter-personalen Beziehungen gegenüber anderen Personen, die bestimmte Wirkungen *verursacht* haben, geübt. Sie beinhaltet einen retributiven oder vergeltenden Akt gegenüber einem personalen Verursacher von Wirkungen auf die Interessen eines Akteurs. Das weist auf drei wesentliche Aspekte hin, die im Falle reziproker Akte vorliegen müssen:

(R. 1) Sie sind Teil eines sequentiellen Prozesses, bei dem ein Akteur zuerst und ein zweiter daran anschließend handeln.

(R. 2) Die Handlungsfolgen der Aktionen des ersten Akteurs können diesem vom zweiten Akteur zugerechnet werden.

(R. 3) Der zweite Akteur wünscht „Gleichheit“, die er in einer Art geistiger Buchhaltung misst, herzustellen.

Nimmt man das Konzept der personalen Beziehung und der retributiven Emotionen, die dabei eine Rolle spielen, ernst, dann wird man verlangen, dass die Handlungen des ersten Akteurs der Handlungssequenz intentional erfolgten. Streng genommen sollten die Handlungen auf eine bewusste Entscheidung des

Erstakteurs zurückzuführen sein. Auch die retributive Handlung, in der sich die Disposition des Zweit-Akteurs zu reziprokem Verhalten ausdrückt, sollte – wiewohl sie durchaus spontan sein kann – aufgrund einer bewussten Entscheidung zustande gekommen sein. Bezieht man diese etwas strengeren Bedingungen ein, dann gelangt man zu einem theoretisch idealisierten Konzept von Reziprozität, welches die folgenden drei Bedingungen voraussetzt.

(R'1) Es liegt ein sequentieller Entscheidungsprozess vor, in dem ein Erst-Akteur zunächst aufgrund einer bewussten Entscheidung positive oder negative Konsequenzen für einen Zweitakteur herbeiführt, der dann seinerseits aufgrund einer bewussten Entscheidung mit positiven oder negativen Sanktionen reagiert.

(R'2) Der Zweitakteur kann den Erstakteur als Verursacher der positiven oder negativen Konsequenzen identifizieren und den intentionalen Charakter der betreffenden Handlungen erkennen, um darauf ebenfalls intentional zu reagieren.

(R' 3) Der zweite Akteur vergilt Positives mit Positivem, Negatives mit Negativem, bis „Gleichheit“ im Sinne seiner „geistigen Buchhaltung“ hergestellt ist.

Die vorangehenden Charakteristika scheinen zum Kern einer sozialtheoretischen Explikation des Reziprozitätskonzeptes zu gehören (zum Explikationskonzept Carnap, 1956, und Siegart, 1997). Wie immer bei zentralen Begriffen kann das gesamte Begriffsfeld jedoch nur über „Familienähnlichkeiten“ bestimmt werden. Eine Kennzeichnung durch ein allgemeines Prädikat, welches notwendig und hinreichend für das Vorliegen des interessierenden Phänomens ist, gibt es nicht.

Wenn etwa Blutspender zur Blutspende motiviert sind, dann wollen sie sich typischerweise fair an einem Prozess beteiligen, von dem sie selbst profitieren, ohne durch ihr Handeln direkt auf einen personalen Verursacher positiver oder negativer Konsequenzen reagieren zu können. Darüber hinaus verschwimmt gerade in derartigen Fällen die Grenze zwischen dem Konzept der Reziprozität und dem des einseitigen Altruismus, der Solidarität oder auch dem Konzept

generalisiert (im Gegensatz zu „personalisiert“) reziproker Beziehungen (vgl. zu indirekter Reziprozität Dufwenberg et al., 2001). Wir werden uns im Weiteren nur mit Fällen befassen, in denen Reziprozität im engeren Sinne des Begriffes vorliegt und somit vom Vorliegen der Bedingungen (R'1), (R'2) und (R'3) ausgehen. Daraus ergibt sich unmittelbar die Notwendigkeit, den intentionalen Aspekt, das Element bewusster Entscheidung zu modellieren. Wie das Bei-Spiel der Ultimatumspiel-Experimente zeigt, scheint dieser Aspekt auch empirisch oder faktisch eine wichtige Rolle zu spielen. Die Individuen gehen in einer Weise vor, die als Ausdruck bewussten Entscheidens aufgrund retributiver Emotionen erscheint und daher auch entsprechend modelliert werden sollte. Damit wird die entwickelte Begrifflichkeit empirisch relevant.

2.2. Das Beispiel des Ultimatumspiels

Das Ultimatumspiel und die an dieses Spiel anschließenden Experimente (vgl. ursprünglich Güth, Schmittberger und Schwarze, 1982, und als spätere Überblicke Güth, 1995, und Roth, 1995) verdeutlichen exemplarisch das bislang zum Methodologischen und Begrifflichen Gesagte. Im Ultimatumspiel wird ein positiver Geldbetrag, p , (der sogenannte „Kuchen“, „pie“) auf zwei Parteien X und Y in Abhängigkeit von deren Entscheidungen aufgeteilt. Dabei gelten die folgenden Regeln:

Zunächst schlägt der Erst-Akteur X eine Aufteilung von p in der Form (x, y) mit $x, y \geq 0$ und $x+y=p$ vor.

Der Zweit-Akteur Y kann den Vorschlag (x, y) annehmen, worauf X den Betrag x erhält und Y den Betrag y , oder ablehnen, worauf beide überhaupt nichts bekommen, d.h. der Auszahlungsvektor ist $(0,0)$.

Wenn man dieses Spiel mit den Mitteln der allgemeinen Rationalwahltheorie analysiert, indem man den Spielern ein allein an den materiellen Auszahlungen

orientiertes opportunistisches rationales Verhalten unterstellt, dann muss ein rationaler Zweit-Akteur im Falle einer einmaligen isolierten Interaktion mit dem Erst-Akteur jede Aufteilung, die ihm einen Wert $y > 0$ bietet, annehmen. Die Erst-Akteure sollten dies antizipieren und nur einen minimalen Betrag y anbieten und die Zweit-Akteure alle derartigen Angebote annehmen. Nimmt man an, dass der reagierende Akteur oder Reakteur nur bei dem hinreichenden Grund, einen echten Verlust zu vermeiden, ablehnt, dann sollte ein Akteur in der zutreffenden Erwartung, dass der Reakteur dieses Angebot annimmt, nur $y=0$ bieten.

Wenn wir derartige Verhaltensweisen relativ selten beobachten (vgl. etwa Güth, Schmidt und Sutter, 2001), dann ist das noch nicht notwendig als Falsifikation des Modells opportunistisch rationalen Verhaltens anzusehen. Insbesondere könnte man darauf verweisen, dass die reagierenden Individuen Reputationseffekte und die Möglichkeit weiterer zukünftiger Interaktionen in ihre Entscheidungen als mögliche künftige Kausalfolgen dieser Entscheidungen einbeziehen (vgl. zu Reputationseffekten Klein, 1997). In spieltheoretischen Experimenten hat man daher sorgfältig darauf geachtet, dass derartige „Rationalisierungsmöglichkeiten“ von direkt unvorteilhaftem Verhalten durch indirekte Vorteile ausscheiden. Man hat Akteure insbesondere anonym und mit der glaubwürdigen Zusicherung der Einmaligkeit der Interaktion das Spiel miteinander spielen lassen.

Insbesondere in jüngerer Zeit durchgeführte Zeitungsexperimente zu Ultimatumspielen müssen den letzten Zweifel daran zerstreuen, dass die beteiligten Probanden auch wirklich darauf vertrauen, niemals mehr mit den gleichen Individuen zu interagieren (vgl. Güth, Schmidt und Sutter, 2002). Sie lösen jede soziale Einbettung der Interaktion für die Beteiligten glaubwürdig auf (vgl. Granovetter, 1985, zum Phänomen der „embeddedness“). Die Ablehnung eines Ultimatums ist unter solchen Bedingungen eindeutig unvorteilhaft, weil jegliche Kausalfolgen einzelner Annahme-Akte auf zukünftige Interaktionen außerhalb des betrachteten Experimentes ausgeschlossen sind. Ein ausschließlich an materiellen Auszahlungen interessierter opportunistischer Zweit-Akteur muss in einer anonymen Einmal-Interaktion deshalb jedes positive Angebot $y > 0$ annehmen. Er hat keinen materiellen Grund, ein Angebot $y=0$ abzulehnen.

Selbst unter Anonymitätsbedingungen werden von den Probanden häufig „faire“ Angebote $(\frac{p}{2}, \frac{p}{2})$ gemacht. Diese werden praktisch immer angenommen. „Unfaire“ Angebote (x, y) mit $x > \frac{2}{3}p$ jedoch, werden selbst bei signifikanten Opportunitätskosten $y \gg 0$ häufig abgelehnt (vgl. Slonim und Roth, 1998, sowie Cameron, 1999).

Der Widerspruch zwischen Spiel-Experiment und Spiel-Theorie ist offenkundig. Jedenfalls dann, wenn man eine Orientierung an materiellen Auszahlungen zugrundelegt, ist die Annahme eines allgemein bekannten Opportunismus nicht mit den Befunden vereinbar. Eine Spezifikation des ökonomischen Verhaltensmodells, die diesen Sachverhalten Rechnung trägt, ohne ad hoc zu sein oder die Theorie letztlich gehaltlos werden zu lassen, erscheint als unerlässlich.

2.3 Erklärungen für die Beobachtungen in Ultimatum-Experimenten

Neo-Klassiker haben in Reaktion auf die skizzierten Befunde gern den Weg einer Modifikation der Nutzenfunktionen beschritten. Sie haben durch unterschiedliche derartige Anpassungen ihre Spielmodelle mit der beobachteten Ablehnung auch positiver Angebote in Ultimatum-Experimenten vereinbar gemacht (vgl. Bolton, 1991, Bolton und Ockenfels, 1999, Fehr & Schmidt, 1999, Kirchsteiger, 1994, und Rabin, 1993). Soweit es dabei in Analogie zum „sophisticated curve fitting“ um „sophisticated game fitting“ geht, ist diese Methode fragwürdig. Das Vorliegen oder Nichtvorliegen von motivationalen Faktoren muss zu weiteren Hypothesen führen, die in ihrem Gehalt über die Datensätze, an denen sie entwickelt wurden, hinausgehen und im Prinzip in unabhängigen weiteren Experimenten überprüft werden können. Wenn das der Fall ist, dann kann man den Vorwurf einer bloßen Ad-hoc-Reparatur vermeiden. Aber auch nur dann.

Die Anreicherung der Nutzenfunktion um zusätzliche motivationale Faktoren bildet vermutlich ein sinnvolles (progressives) Forschungsprogramm. Die Faktoren müssen sich allerdings, um es klassisch zu formulieren, in ein

Gesamtbild von der „menschlichen Natur“ einfügen, das sowohl die Fähigkeiten zu opportunistischer Rationalität als auch die Verhaltensbestimmung nach internalisierten Normen und Werten zulässt. Im vorliegenden Falle bedeutet das letztlich, dass das Auftreten motivationaler Faktoren oder „Präferenzen“ selbst einer Erklärung zu unterziehen ist. Benutzen wir, was im Kontext einer Diskussion um die menschliche Natur ohnehin nicht fern liegt, die Terminologie der Evolutionstheorie, dann geht es nicht nur um eine „proximate“ Erklärung der Daten, sondern um eine „ultimate“ Erklärung dafür, warum und wie der unterstellte erklärende Faktor entstehen und sich in einer konkurrenzbestimmten Interaktion halten konnte.³

Letztlich müssten adaptive Prozesse angegeben werden, in denen sich die betreffenden Anlagen herausgebildet haben. Die Kausalmechanismen, die der Entwicklung von bestimmten Verhaltensdispositionen, angeborenen wie erlernten Faktoren zu Grunde liegen, bedürften an sich der genauen empirischen Aufklärung. Das ist ein ganzes Forschungsprogramm, das wir weder hier noch an anderer Stelle abarbeiten können. Wir wollen jedoch einen konzeptionellen Beitrag leisten, um aufzuzeigen, wie man im Prinzip verschiedene Verhaltensansätze miteinander verbinden kann.

Der indirekte evolutionäre Ansatz, den wir nun in Anwendung auf das Reziprozitätsproblem skizzieren wollen, erlaubt es, von zukünftigen Erwartungen bestimmtes, opportunistisch rationales und von vergangenen Erfahrungen geprägtes, gleichsam „programmiertes“ Verhalten systematisch im gleichen Modell zu berücksichtigen. Der Ansatz bietet, wie wir meinen, letztlich Raum für beides, die klassischen „intentionalistischen“ Erklärungen der reinen ökonomischen Entscheidungstheorie und die „nicht-intentionalistischen“ natur- und sozialwissenschaftlichen Erklärungen etwa von Soziologie und Psychologie.

Gegenwärtig haben wir allerdings nur das weit bescheidenere Ziel, die potentielle Fruchtbarkeit des Ansatzes an dem konkreten Beispiel einer im weiteren Sinne ökonomischen Behandlung von Reziprozitätsmotiven zu illustrieren.

³ Die proximate Erklärung dafür, warum Schneehasen weiß sind, ist die Beschaffenheit ihres Felles zusammen mit den Gesetzen der Optik, die ultimate der Selektionsdruck jagender Polarfüchse in einer überwiegend hellen Umwelt zusammen mit grundlegenden evolutionsbiologischen Gesetzen.

Wir halten die Indizien für das Vorliegen einer Reziprozitätsneigung grundsätzlich durchaus für bedenkenswert. Reziprozität modellieren wir im Anschluß an einige der vorerwähnten jüngeren Studien über eine Ungleichheitsaversion. Im Unterschied zu diesen Studien gehen wir jedoch davon aus, dass diese Aversion im Sinne unserer Reziprozitätsdefinition ausschließlich auf Individuen in der Rolle als Reakteur wirkt. Der Reakteur unterliegt retributiven Emotionen, die von vorangehenden Handlungen des Akteurs und nicht allein von einer ungleichen Verteilung von Ergebnissen oder Resultaten ausgelöst werden.

Entstehende Ungleichheit wird vom Reakteur dem Akteur zugerechnet. Es geht also nicht nur darum, ob Ungleichheit der Ergebnisse vorliegt, sondern darum, wie die Ungleichheit zustande kam.

3. Ein indirekt evolutionärer Zugang zum Reziprozitätsproblem

Alle evolutionären Ansätze verknüpfen beobachtbares Verhalten mit der Vergangenheit. Für den indirekt evolutionären ist im Unterschied zum traditionellen direkt evolutionären Ansatz ausschlaggebend, dass er nicht nur die (subjektive) Entscheidungssituation des Entscheiders explizit modelliert, sondern, dass er Komponenten zulässt, die zwar das Entscheidungsverhalten beeinflussen, jedoch keinen direkten evolutionären Bezug haben. „Objektive“ Auszahlungen – in der klassischen Evolutionstheorie sind dies Anteile am Genpool oder Anzahl der Nachkommen – bestimmen, was sich in einem evolutionären Prozess letztlich durchsetzt. In der direkten evolutionären Analyse wird jede Aktion eines Akteurs direkt nach den objektiven Auszahlungen als vorteilhaft oder unvorteilhaft bewertet. Evolutionär setzt sich das durch, was relativ erfolgreicher ist. Es kommt also auf die Relationen zwischen den objektiven Auszahlungen an.

Das ist zwar in der indirekten evolutionären Analyse am Ende nicht anders, sonst wäre sie keine evolutionäre Analyse, doch bietet die indirekte Analyse auf eine neue Weise Raum für Entscheidungsprozesse, die zu Handlungen führen,

die „direkt“ unvorteilhaft scheinen.⁴ Im indirekten evolutionären Ansatz wird der Einfluss opportunistisch rationalen Entscheidens auf das Verhalten als Teil eines evolutionären Modells erfassbar. Zukunftsgerichtetes rationales Entscheiden wird zugelassen. Das Verhalten wirkt sich möglicherweise in direkt unvorteilhafter Weise auf die objektiven Auszahlungen aus. Die Motive des Entscheiders lassen ihm dieses Verhalten frei nach Nietzsche aber aufgrund einer „Umwertung der Werte“ subjektiv als bevorzugenswert erscheinen. Das Vorliegen der subjektiven Motive kann generell evolutionär vorteilhaft sein, obschon die Motive im Einzelfall zu direkt unvorteilhaftem Verhalten führen.

Neo-klassische Ökonomen müssen für das Auftreten direkt unvorteilhaften Verhaltens eine Erklärung im Einklang mit neo-klassischen Prinzipien finden. Sie befinden sich auf dem Holzwege, wenn sie meinen, eine neo-klassische Erklärung für das beobachtete Verhalten nach generellen Dispositionen oder Präferenzen liege bereits vor, wenn sie aufzeigen können, dass die Dispositionen für deren Besitzer materiell vorteilhaft sind. Nach dem Modell opportunistischen (rationalen) Verhaltens zählen die Kausalfolgen des einzelnen Aktes. Ob ein Akt einer Klasse von Handlungen zugehört, deren allgemeine Durchführung wünschenswerte Folgen für den Besitzer *hätte*, ist im neo-klassischen Ansatz nicht relevant. Ein opportunistisches rationales Individuum rechnet jeder Handlung genau die von dieser Handlung ausgehenden erwarteten Kausalfolgen zu und bewertet jede Entscheidung je für sich nach diesen Kausalfolgen der Einzelentscheidung.

Ein rationaler Akteur kann sich daher zwar vornehmen, eine Sequenz von Handlungen durchzuführen. Eine Strategie als *Plan* kann er in dem Sinne wählen, in dem man sich für einen Plan entscheiden kann. Wenn man opportunitäten-ergreifendes Rationalverhalten zu Grunde legt, muss aber in *Durchführung* des Planes jede Entscheidung je für sich (sequentiell) im Lichte der Kausalfolgen der Einzelentscheidung getroffen werden. Man kann erst handeln, während man den Plan durchführt und dann steht man immer vor

⁴ Im evolutionären Modell finden Verhaltensweisen, die direkt nur anderen nützen, Verhaltensweisen, die direkt unvorteilhaft für den Akteur sind oder die von Präferenzen und Emotionen getragen sind, die dem direkten Eigeninteresse des von diesen Faktoren geleiteten Akteurs widersprechen, traditionell nur über den Umweg über Verwandtschaftsselektion und ähnliche komplexe Mechanismen Platz.

neuen Opportunitäten (und auf diese zu beziehenden Opportunitätskosten). Für Verhaltensdispositionen oder die indirekt vorteilhafte Anlage, auf direkte Vorteile zu verzichten, ist im traditionellen Ansatz zunächst kein Platz, wohl aber in einem indirekten evolutionären Ansatz. Wenden wir uns diesem zu, indem wir seinen Beitrag zur konkreten Behandlung des Reziprozitätsthemas und der retributiven Neigungen untersuchen.

3.1. Der indirekt evolutionäre Zugang zum Ultimatumspiel

Die vorangehenden allgemeinen Überlegungen lassen sich konkret am Ultimatumspiel und einer „Reziprozitätsneigung“, die die subjektive Bewertung objektiver Auszahlungen so beeinflusst, dass der Zweitakteur oder „Reakteur“ auch positive objektive Auszahlungen zurückweist, illustrieren. In einem indirekt evolutionären Modell zum Ultimatumspiel führt man beispielsweise Nutzenfunktionen der folgenden Art für den Reakteur ein:

$$(1) \quad U_Y(x, y) = y - i \cdot \max\{0, x - y\}, \quad i \geq 0$$

Dabei ist i ein Parameter, der die Entscheidungen via $-i \cdot \max\{0, x - y\}$ beeinflusst, ohne selbst direkten materiellen Auszahlungsbezug aufzuweisen. Die subjektive Bewertung eines objektiven Angebotes y wird durch die „reaktive Ungleichheitsaversion“ reduziert zu $y - i \cdot \max\{0, x - y\}$. Der Reakteur Y sieht nicht primär „prospektiv“, dass er durch Annahme des Angebotes y anstatt 0 bekommen würde. Er sieht vielmehr „retrospektiv“, dass er weniger als die Hälfte von p zugeteilt erhielt und sich durch Annahme des Angebotes mit dieser Behandlung durch X zufrieden geben würde. Er reagiert auf ein Angebot von $y < \frac{1}{2}p$ mit einer retributiven oder vergeltenden „Emotion“. Die subjektive Bewertung der objektiven Konsequenzen der Annahme des Angebotes werden durch den „Frustrationsfaktor“ $i(x - y)$ zu $y - i(x - y) \leq y$. Lehnt Y das Angebot ab, so kann er seinen Emotionen Raum geben und bekommt keinerlei objektive

Auszahlungen, nimmt er an, so ist ihm der subjektive Nutzen des objektiven Betrages y zumindest um $i(x-y)$ vergällt.

Der Faktor $i(x-y)$ wird nach der hier vertretenen Sicht der Dinge also davon beeinflusst, wie das Ergebnis entstand. Die subjektive Bewertung der Spielergebnisse ist insoweit von der Spielstruktur und nicht nur von der Ergebnisverteilung abhängig. Es geht nicht um eine einfache Aversion gegen ungleiche Ergebnisse, sondern um eine komplexere emotionale Reaktion, die zu einer „subjektiven Umwertung der objektiven Konsequenzen“ führt.

Das Verhältnis von objektiver und subjektiver Ebene versteht man besser, wenn man sich klar macht, dass zwei Individuen mit unterschiedlichem i , dann, wenn sie die gleichen Entscheidungen treffen, auch die gleichen objektiven Auszahlungen erhalten – obschon sie diese subjektiv unterschiedlich bewerten. Zwei Individuen Y' , Y'' , mit $i' < i''$ entgeht bei einem gegebenem Angebot $y > 0$, das sie ablehnen, in materiellen Auszahlungen jeweils $y > 0$. Wenn $0 > U_{Y'}(x, y) > U_{Y''}(x, y)$ gilt, handeln beide als Maximierer ihres subjektiven Erwartungsnutzens rational, wenn sie die positive materielle Auszahlung $y > 0$ zurückweisen. Wenn $U_{Y'}(x, y) > 0 > U_{Y''}(x, y)$ gilt, dann sollte Y' aufgrund seiner subjektiven, von i' abhängigen Bewertung das Angebot annehmen und Y'' es ablehnen. In diesem Falle werden die Akteure aufgrund der unterschiedlichen subjektiven Parameter i' , i'' unterschiedlich handeln und auch unterschiedliche objektive Auszahlungen von $y > 0$ bzw. 0 erhalten. Sollte allerdings bspw. der Akteur Y'' einen Fehler machen und ungeachtet von $U_{Y''}(x, y) < 0$ das Angebot annehmen, dann erhielte er wiederum die gleiche objektive Auszahlung y wie der Akteur Y' , obschon er subjektiv das Ergebnis als Nutzenminderung betrachten würde. Im dritten Fall $U_{Y'}(x, y) > U_{Y''}(x, y) > 0$ würden beide annehmen und damit ein äußerlich gleiches Verhalten mit gleichen objektiven Konsequenzen aufgrund unterschiedlicher subjektiver Bewertungen zeigen.

Aufgrund seiner subjektiven Bewertungen wird ein Reakteur Y alle Angebote annehmen, die $y \geq \frac{i}{1+2i}p$ erfüllen, während er alle anderen ablehnt. Das Ergebnis einer Interaktion, in der Reziprozität des Zweitakteurs eine Rolle spielt, hängt wesentlich davon ab, ob und inwieweit die Stärke der intrinsischen Motivation zur Ablehnung einer Ungleichbehandlung und damit die Höhe von i bekannt ist. Nimmt man an, dass i gemeinsames Wissen ist, so wird X aufgrund

seiner Kenntnis von i dazu gebracht, für jeden Kuchen p und jeden Wert von i genau das Annahmeminimum

$$(2) \quad y^*(i) = \frac{i}{1+2i} p$$

anzubieten.

Individuen, die ein höheres i aufweisen, werden dann, wenn i gemeinsames Wissen ist, besser abschneiden als Individuen, mit niedrigerem i . Denn die Angebote y , die gerade noch angenommen werden, müssen mit größerem i immer näher bei $\frac{p}{2}$ liegen.

Ist hingegen die Höhe von i private Information des Reakteurs Y , so ist es niemals schlechter für Y , einen kleineren Wert von i aufzuweisen, und manchmal besser. Ein Reakteur mit kleinerem i nimmt noch Angebote $y > 0$ an, die ein Reakteur mit größerem i ablehnt. Wenn wir unterstellen, dass das Auftreten von Angeboten $y > 0$ etwa aufgrund von Fehlern bei der Strategiewahl niemals vernachlässigt werden kann, schneiden die Reakteure mit kleinerem i im Schnitt also besser ab, weil sie alle Angebote annehmen können, die die Reakteure mit größerem i annehmen würden und zusätzlich noch einige weitere, deren Annahme Reakteuren mit größerem i verschlossen bleibt. Ein höheres i kann aber keinen evolutionären Vorteil bieten, wenn i private Information von Y ist. Denn der Akteur X kann mangels Kenntnis von i nicht dazu gebracht werden, höhere „objektive“ Angebote $y > 0$ zu machen, um eine Ablehnung durch Y zu vermeiden.

Die bisherige Diskussion betrifft extreme Informationsbedingungen. Entweder ist der Parameter i gemeinsames Wissen oder er ist ausschließlich seinem Träger bekannt. Dauerhafte Kleingruppenbeziehungen, in denen der Parameter i einer Person mit einer gewissen Plausibilität vom privaten zum gemeinsamen Wissen werden kann, sollten nach den vorangehenden Ergebnissen andere Eigenschaften als weitgehend anonyme Interaktionen etwa auf großen Märkten mit weitgehend privater Information über i haben. Es ist allerdings in sich eine etwas merkwürdige Annahme, dass der i -Wert von Personen entweder

vollkommen bekannt oder vollkommen unbekannt sein muss. Wir Menschen machen de facto Erfahrungen mit einer Vielzahl von Partnern in unserer Gesellschaft. Wir sind in einer bestimmten Weise sozialisiert worden und kennen daher introspektiv zumindest unsere eigene Neigung, uns gegen ungleiche Behandlung, die wir als unfair bewerten, auch unter Kosten für uns selbst zu wehren. Wir haben nicht nur über spezielle Partner, sondern auch über das allgemeine „soziale Klima“ in der Population unserer potentiellen Partner Informationen. De facto werden daher in der Regel nicht die beiden extremen, sondern eher „mittlere“ Informationsbedingungen vorliegen, die sowohl allgemeine Kenntnisse über die Zusammensetzung der Population möglicher Partner als auch (potentiell kostenträchtige) Möglichkeiten der Informationsbeschaffung über einen speziellen Partner beinhalten. – Wir wollen das bislang vorgestellte Modell nun in dieser Weise ergänzen (vgl. zu einem analogen Vorgehen Güth und Kliemt, 2000, zur Evolution von Vertrauenswürdigkeit im Falle des Vertrauensspieles).

3.2. Imperfekte Partner-Information

Zur Vereinfachung sei angenommen, der Parameter i könne nur die beiden Werte $i=0$ und $i=1$ annehmen. Entweder betrachtet der Reakteur die Interaktion im Bezugsrahmen fairer Kooperation und läßt sich daher von einer Reziprozitätsneigung leiten, die sich mit der Stärke $i=1$ in eine Senkung seines subjektiven Nutzens ungleicher Aufteilungen umsetzt, oder der Partner sieht die Interaktion allein unter dem Aspekt materieller Auszahlungsmaximierung. Anders ausgedrückt, jedes Individuum aus der Menge möglicher Partner ist entweder mit der „Stärke“ $i=1$ „intrinsisch motiviert“, retributiv die Herstellung „fairer Gleichheit“ auch gegen die eigenen objektiven materiellen Interessen zu betreiben, oder das Individuum ist ausschließlich „extrinsisch motiviert“, nach einer Verbesserung seiner je eigenen objektiven Lage zu streben.

Unter diesen vereinfachenden Annahmen hat man es mit einer Population entweder (partiell) intrinsisch, $i=1$, oder extrinsisch, $i=0$, motivierter Partner zu tun. Es sei, um ein für unsere illustrativen Zwecke hinreichendes einfaches

evolutionäres Modell zu erhalten, weiterhin unterstellt, dass Paare von Individuen aus einer unendlichen Population (Bedingungen endlicher Populationen werden diskutiert in Güth et al., 2000) zufällig gebildet werden, um jeweils Ultimatumspiele zu spielen, ohne dabei die Erwartung zu hegen, erneut auf den gleichen Partner zu treffen. Die objektiven Auszahlungen in diesen Spielen messen den Erfolg intrinsisch wie extrinsisch motivierter Individuen.

Der relative objektive Erfolg dieser beiden Spieler-Typen bestimmt im Zuge der Evolution über die Populationszusammensetzung, die durch den Anteil a , $a \in [0, 1]$, intrinsisch motivierter Individuen charakterisiert wird. Wenn die objektiven erwarteten Auszahlungen der intrinsisch motivierten $i=1$ -Typen oder kurz, der 1-Typen, höher sind als die objektiven erwarteten Auszahlungen der ausschließlich extrinsisch motivierten 0-Typen, wird der Anteil der 1-Typen an der Population bzw. die Variable a wachsen. Wenn die objektiven Auszahlungen der 0-Typen im Schnitt höher als die der 1-Typen sind, wird a fallen.⁵ Es wird angenommen, dass a als Variable, die die Gesamtpopulation charakterisiert, gemeinsames Wissen der Akteure ist.

Wenn die Spieler X und Y zufällig ausgewählt werden, um ein Ultimatumspiel miteinander zu spielen, dann bestimmt nach den bisherigen Annahmen der Wert a die A priori-Wahrscheinlichkeit a , mit der der Erstakteur X erwartet, auf einen intrinsisch motivierten Zweitakteur Y zu treffen. Wir nehmen nun zusätzlich an, dass X die Option erhält, weitere Informationen über den i -Typ von Y für Kosten von $C > 0$ einzuholen. Grundsätzlich kann es sich dabei um ein Signal von begrenzter Zuverlässigkeit handeln, jedoch setzen wir zur Vereinfachung voraus, dass das Signal dem Erstakteur X den Typ des Zweitakteurs Y zuverlässig (mit Sicherheit) offenbart (wobei im Falle eines stochastischen Signals nahezu alle im weiteren gemachten Aussagen grundsätzlich in ähnlicher Weise, wenn auch technisch etwas komplizierter abzuleiten sind; vgl. wiederum Güth und Kliemt, 2000).

Unter diesen Prämissen läßt sich bestimmen, bei welchen Konstellationen der übrigen Parameter der Erstakteur X die Kosten C für die Information über Y

⁵ Da wir nur generische Fälle betrachten, ist hier nicht näher auf den Fall der Auszahlungsgleichheit einzugehen.

aufbringen sollte. Aus Sicht von X sind nur die beiden jeweils optimalen Aufteilungsvorschläge $y^*(i) \mid_{i=0}=0$, falls er es mit einem 0-Typ zu tun hat, und $y^*(i) \mid_{i=1}=p/3$, falls er es mit einem 1-Typ zu tun hat, zu berücksichtigen. Ein rationaler Entscheider X wird eine der beiden Aufteilungen wählen. Welche er wählen wird, hängt von den Auszahlungserwartungen ab, die er mit den Aufteilungen jeweils verknüpft. Die Erwartungen wiederum hängen davon ab, welche Informationen der Aufteiler X über den Typ des Reakteurs Y besitzt. Der Aufteiler kann sich hier entscheiden, entweder aufgrund der A priori-Information a zu handeln oder zusätzliche Informationen einzuholen. Ob er die Kosten C für das kostenpflichtige Signal aufbringen soll, hängt davon ab, ob die so erreichbare Steigerung der Auszahlungserwartung größer ist als die Kosten C .

Bei gegebener a priori Wahrscheinlichkeit $a \in [0, 1]$ für einen 1-Typ beträgt die erwartete Auszahlung, falls er $y=0$ zuteilt, $a \cdot 0 + (1-a) \cdot p$ und $(2/3) \cdot p$, falls er $y=p/3$ zuteilt. Das erstere sollte er genau dann wählen, wenn $(1-a) \cdot p > (2/3) \cdot p$. Damit ist – vorausgesetzt, dass X das kostenpflichtige Signal nicht kauft – das für den Erstakteur optimale Angebot

$$(3) \quad y^*(a) = \begin{cases} y = 0 & \text{für } a < \frac{1}{3}, \\ y = \frac{p}{3} & \text{sonst.} \end{cases}$$

Erwirbt X die Information über den i-Typ, so wird er eine Aufteilung mit $y=0$ vorschlagen, nachdem ein 0-Typ und $y=p/3$, nachdem ein 1-Typ signalisiert wurde. Im ersten Falle, der mit Wahrscheinlichkeit $(1-a)$ eintritt, beträgt die Erwartung $(1-a)p$, im zweiten Falle, der mit Wahrscheinlichkeit a auftritt, beträgt die Auszahlungserwartung hingegen $\frac{2}{3}ap$. Bringt man davon die Kosten C , die mit Sicherheit aufgewandt werden müssen, in Abzug, so ergibt sich eine Gesamterwartung von $(1-a)p + \frac{2}{3}ap - C$.

Für $a \geq 1/3$ lohnt sich der Ankauf von spezifischer Typeninformation somit dann, wenn gilt

$$(3) \quad (1-a)p + \frac{2}{3}ap - C \geq \frac{2}{3}p \quad \Leftrightarrow \quad \frac{1}{3}(1-a) \geq \frac{C}{p};$$

während sich für $a < \frac{1}{3}$ der Ankauf von spezieller Typeninformation dann lohnt, wenn gilt

$$(4) \quad (1-a)p + \frac{2}{3}ap - C \geq (1-a)p \quad \Leftrightarrow \quad \frac{2}{3}a \geq \frac{C}{p}.$$

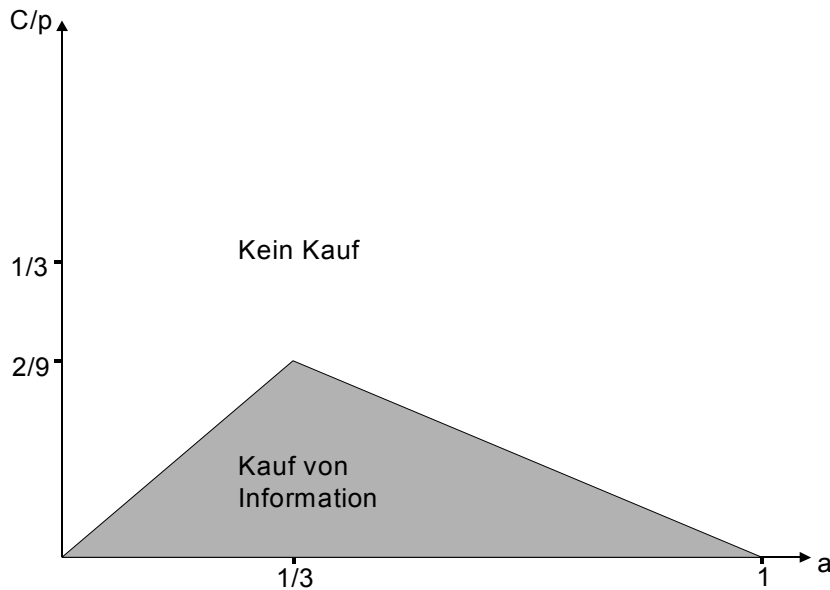


Abbildung 1: Erwerb von Typ-Information im Ultimatumspiel

Unter den gemachten Annahmen reiner Zufallspaarung von Partnern in Ultimatumsspielen der beschriebenen Art werden alle rationalen Akteure X spezifische Typeninformation zu Kosten von C erwerben, falls die Parameterkombination $(a, \frac{C}{p})$ in dem in Abbildung 1 skizzierten Dreieck liegt, während keine spezielle Typeninformation erworben wird, wenn die Kombination außerhalb des Dreiecks liegt.

Die Auswirkungen dieser Sachverhalte auf die Populationsdynamik sind offenkundig. Wenn $(a, \frac{C}{p})$ außerhalb des Dreiecks liegt, so wird über die A priori-Information hinaus kein Erstakteur zusätzliche Typeninformation erwerben. Das führt uns auf die Diskussion im vorangehenden Abschnitt zurück. Dort wurde der Fall rein privater Information über den i -Typ des Reakteurs betrachtet. Im Falle privater Information war es ein Nachteil, einen positiven i -

Wert aufzuweisen. Einen Unterschied zum vorangehenden Abschnitt stellt lediglich unsere Einschränkung $i \in \{0,1\}$ dar. Wenn $a < \frac{1}{3}$ gilt, so schneiden grundsätzlich beide Typen in der Rolle des Zweitakteurs gleich ab. Nur dann, wenn der Erstakteur einen Fehler machen sollte, kann es einen echten Vorteil beinhalten, zur opportunistisch rationalen Annahme des Angebotes in der Rolle des Zweitakteurs befähigt zu sein. Es kommt also in dem Bereich außerhalb des Dreiecks, der zugleich links von $a = \frac{1}{3}$ liegt, zu einer, wenn auch nur schwachen, von gelegentlichen „trembles“ bestimmten Tendenz zur Abnahme des Anteils a der 1-Typen.⁶ Für $a > \frac{1}{3}$ werden außerhalb des Dreiecks aufgrund der A priori-Information unabhängig von deren Typ allen Zweitakteuren von rationalen Erstakteuren positive Angebote von $p/3$ gemacht, die von beiden Typen angenommen werden. Es kommt dadurch in diesem Bereich ebenfalls nur zu einer schwachen, aber im Falle von seltenen „Angebotsfehlern“ (im Sinne von fehlerhaften Angeboten $y < p/3$) systematischen Tendenz zur Abnahme der 1-Typen. Unterhalb der Kurve wirkt sich die Kenntnis des i -Typs des Reakteurs zugunsten der Typen mit dem höheren i aus. Die Zweitakteure vom 1-Typ schneiden besser ab, weil ihnen $p/3$ angeboten und dieses von ihnen angenommen wird, während die Zweitakteure vom 0-Typ entdeckt werden und nichts oder nur einen minimalen Betrag angeboten erhalten.

Das nachfolgende Pfeildiagramm zeigt die schwachen Abnahmetendenzen außerhalb und die starken Zunahmetendenzen innerhalb des Dreiecks auf. Es ist offenkundig, dass die beiden Kandidaten für eine evolutionär stabile Populationszusammensetzung $a=0$ für Anfangskonstellationen $(a, \frac{C}{p})$ außerhalb und $a=1-3\frac{C}{p}$ für Anfangskonstellationen innerhalb des gepunkteten Trapezes sind.

⁶ Konkret ließe sich annehmen, dass der Erst-Akteur mit der Wahrscheinlichkeit $1-\varepsilon$ das optimale Angebot $y^*(a)$ macht, jedoch mit einer kleinen Wahrscheinlichkeit $\varepsilon > 0$ zufällig ein „fehlerhaftes“ Angebot $y \in [0, 1/2]$ unterbreitet; vgl. zum „trembling“ in evolutionären Kontexten vor allem Selten (1983 und 1988).

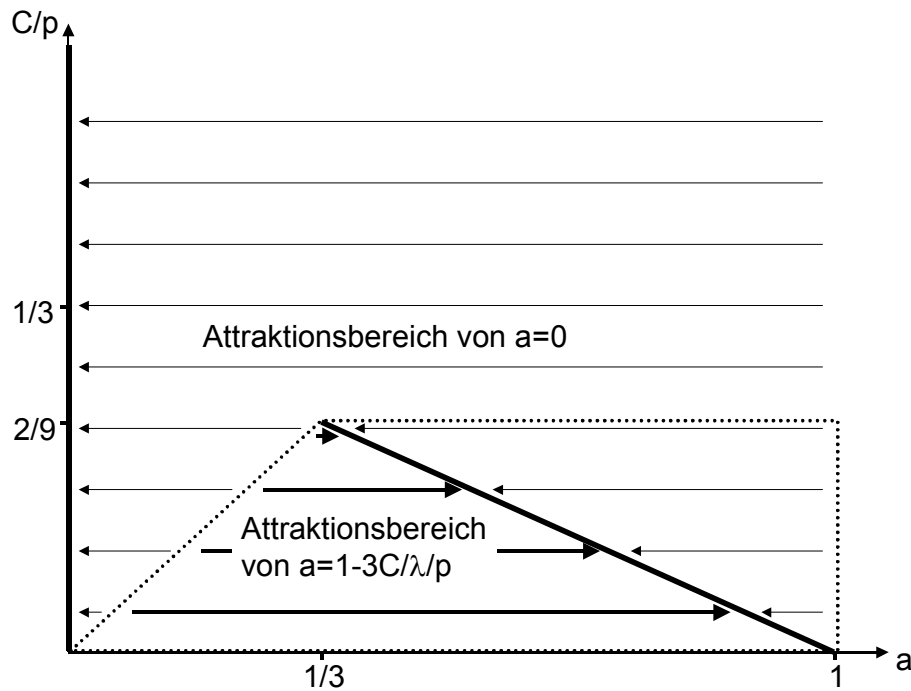


Abbildung 2: Dynamik des Populationsanteils a

Wenn man von der idealisierenden Annahme vollkommen zuverlässig zwischen Typen diskriminierender Signale abrückt und annimmt, dass Erstakteure nur Zugang zu einem stochastischen, aber informativen Signal erwerben können, dann verkleinert sich das Dreieck in Abbildung 2. Die Höhe am Punkt $a=1/3$ auf der Abzisse wird sich verringern und die Eckpunkte des Dreiecks zugleich auf $a=1/3$ zu rücken. Der Attraktionsbereich des evolutionär stabilen Bimorphismus mit $a=1-3(C/p)$ wird sich daher verkleinern. Für dieses Gebiet bietet der indirekt evolutionäre Ansatz mit dem hier angegebenen Modell eine potentielle Erklärung dafür, dass wir in Populationen typischerweise immer „gemischte“ Verhaltensweisen finden. Neben einer evolutionären Nische für normkonformes gibt es immer eine Nische für normabweichendes Verhalten (vgl. Frank, 1987 und 1992, für grundlegende intuitive Überlegungen dieser Art).

3.3. Evolution für mehr als ein Spiel

In den beiden vorherigen Abschnitten haben wir die indirekte Evolution von Reziprozität in einem „Habitat“ betrachtet, welches ausschließlich aus dem Ultimatumspiel besteht. Bei vollständiger Informiertheit des Erst-Akteurs bzw. hinreichend niedrigen Kosten der Informationsbeschaffung, die z.B. für Interaktion in kleinen Gruppen plausibel erscheinen, erweist es sich in dieser Umwelt als ein evolutionärer Vorteil, eine Präferenz für Reziprozität zu besitzen. In diesem Abschnitt soll ergänzend betrachtet werden, wie sich das Ergebnis ändert und die Methodik indirekter Evolution anzuwenden ist, wenn die evolutionäre Umgebung durch mehr als einen Spieltyp definiert ist.

Konkret soll das Habitat die beiden Akteure X und Y mit Wahrscheinlichkeit $\lambda \in (0,1)$ mit dem Ultimatumspiel und mit der Restwahrscheinlichkeit $1-\lambda$ mit dem Diktatorspiel konfrontieren. Letzteres unterscheidet sich vom Ultimatumspiel dadurch, dass Akteur Y jede Aufteilung annehmen *mus*s also nicht mehr auf ein Angebot reagieren kann, sondern rein passiv von Akteur X den Anteil y zugeteilt bekommt. Es stellt sich in dieser aus verschiedenen Spieltypen bestehenden Umwelt wiederum die Frage, welcher i -Wert evolvieren wird, wenn der reproduktive Erfolg allein durch den materiellen Gewinn bestimmt ist.

Ist der i -Wert des Zweit-„Akteurs“ Y dem Erst-Akteur X bekannt, so wird X im Ultimatumspiel wiederum $y^*(i) = \frac{i}{1+2i}p$ anbieten und im Diktatorspiel unabhängig von i die Aufteilung $(x,y)=(1,0)$ bestimmen. Ein großes i ist damit für Y im Ultimatumspiel förderlich und im Diktatorspiel nicht schädlich. Da der Anteil des Ultimatumspiels, λ , positiv ist, bleibt damit das Ergebnis eines im Evolutionsprozess unbeschränkt wachsenden Reziprozitäts-Parameters i (vgl. Abschnitt 3.1) bestehen.

Kennt nur Y seinen i -Wert, so kann dieser, wie in Abschnitt 3.1 diskutiert, das Angebot von X weder im Ultimatum- noch im Diktatorspiel beeinflussen. Damit wird i im Verlaufe des Evolutionsprozesses in einer solchen gemischten Umgebung ebenfalls gegen Null konvergieren.

Was ergibt sich, wenn analog zum vorigen Abschnitt 3.2 Spieler X Information über den i -Wert des Y mit $i \in \{0,1\}$ zum Preis von C beschaffen kann? Es sei dabei unterstellt, dass X über den Informationskauf entscheiden muss, *bevor* er den Spieltyp erkennt. Im Ultimatumspiel ohne Informationskauf gilt weiterhin

$$y^*(a) = \begin{cases} 0 & \text{für } a < 1/3, \\ p/3 & \text{sonst.} \end{cases}$$

Informationskauf lohnt daher für das kombinierte Spiel und $a < 1/3$, falls

$$\lambda[(1-a)p + 2/3ap] + (1-\lambda)p - C \geq \lambda(1-a) + (1-\lambda)p$$

bzw.

$$(5) \quad \frac{2}{3}a \geq \frac{C/\lambda}{p}.$$

Für $a \geq 1/3$ ist die analoge Bedingung

$$\lambda[(1-a)p + 2/3ap] + (1-\lambda)p - C \geq 2/3\lambda p + (1-\lambda)p$$

bzw.

$$(6) \quad \frac{1}{3}(1-a) \geq \frac{C/\lambda}{p}.$$

Der Vergleich von (5) und (6) mit (3) und (4) zeigt den Einfluss des Anteils λ der Ultimatumspiele: Je kleiner er ist, desto stärker schlagen die Kosten C der im Diktatorspiel wertlosen Information zu Buche. Ein Kauf rechnet sich damit für $\lambda \rightarrow 0$ zunehmend weniger, was auf eine Gleichbehandlung aller Zweitakteure Y im Ultimatumspiel hinausläuft und je nach Populationsanteil a entweder durchgängig „faire“ oder aber den Diktatorspiel-Aufteilungen entsprechende Angebote zur Folge hat. Die Attraktionsbereiche der evolutionär stabilen Populationsanteile $a=0$ bzw. $a=1-3\frac{C/\lambda}{p}$ von im Sinne von $i=1$ reziprok eingestellten Individuen entsprechen denen in Abbildung 2 – jedoch mit einer um den Faktor λ gestauchten Ordinatenachse.

Mit zunehmendem λ vergrößert sich der Attraktionsbereich der stabilen Bimorphismen $a \geq 1/3$, während sich die Chancen für derartige Polymorphismen

für $\lambda \rightarrow 0$ verschlechtern. Welche Informationskosten C/p für Reziprozität in dem Sinne prohibitiv wirken, dass für *alle* Startwerte $a \in [0,1]$ gilt $a \rightarrow 0$, hängt linear (in der Form $C/p \geq \frac{2}{3}\lambda$) von λ ab. Es entscheiden also letztlich die Wahrscheinlichkeit von reziprozitätsförderlichen Spieltypen und die relativen Kosten C/p darüber, ob und welche Bimorphismen bezüglich reziproker Veranlagung sich als stabil erweisen.

4. Diskussion

Das Ergebnis der vorangehenden Überlegungen ist intuitiv einleuchtend. Retributive Emotionen, Neigungen dazu, Gutes mit Gutem und Schlechtes mit Schlechtem *reaktiv* zu verknüpfen, können sich unter Konkurrenz-Bedingungen, in denen objektive Erfolgsmaße eine Rolle spielen, nur halten, wenn sie einen positiven adaptiven Wert besitzen. Es wurde im Vorangehenden nichts über den konkreten Mechanismus gesagt, durch den Abweichungen von den Annahmen opportunistischer Rationalität zustande kommen und sich halten können. Ob es sich im Falle der Reziprozitäts-Neigung um eine genetisch verankerte und durch Fortpflanzung weitergegebene Anlage handelt oder ob diese Anlage in sozialen Trainingsprozessen von Generation zu Generation weitergegeben wird, ob wir es eher mit „nature“ oder mit „nurture“ zu tun haben, ist eine bedeutsame Frage, über die wir hier nur spekulieren könnten. Zu ihr geben die Experimente, auf die wir uns bezogen haben – und allgemein ökonomische Experimente etwa im Gegensatz zur Zwillingsforschung – eher wenig her.

Diese Frage kann allerdings dafür, ob es sinnvoll ist, einen indirekten evolutionären Ansatz als allgemeinen Rahmen zur Organisation von „Theorie und Erfahrung“ in der Ökonomik zu benutzen, offen bleiben. Denn solange man den Unterschied zwischen, der Ebene, auf der sich die „Spielregeln“ (zu denen spieltheoretisch ja alle Aspekte des Spieles jenseits des strategischen Einflusses der Spieler und damit auch die Präferenzen der Spieler gehören) entwickeln, und der Ebene der Entscheidungen im Rahmen bestehender Spielregeln machen kann, liegt es prinzipiell nahe, einen indirekten evolutionären Ansatz zu verfolgen.

Dieser Ansatz repräsentiert von seiner inneren Struktur her ein Zwei-Ebenen-Modell, das in der politischen Ökonomie durchgängig immer wieder aufzutreten scheint. Im normativen oder besser „technologischen“ Bereich gehört hierhin die Unterscheidung von „rule and within rule choices“ (Brennan and Buchanan, 1985) und damit die gesamte Konzeption der konstitutionellen politischen Ökonomik und der deutschen Ordnungstheorie bis hin zu v. Hayeks sozialem Evolutionismus (vgl. Vanberg, 1981). Auch hier können experimentelle Studien gewisse Auskünfte über institutionelle Mechanismen und deren Potential geben. Für die im vorliegenden Aufsatz behandelten Fragen ist der Bezug zu anderen Bereichen eher explanativer Natur jedoch bedeutsamer. Denn zwei ganz unterschiedliche Formen von Erklärung, die sich auf ganz unterschiedliche Gesetzestypen zu beziehen scheinen, spielen in der Ökonomik eine Rolle. Diese scheinen sich wiederum nahtlos in ein Zwei-Ebenen-Modell einzufügen.

Zum einen kennt die Ökonomik traditionell „Erklärungen“, die auf allgemeine Systemeigenschaften eines Interaktionssystems abstellen, zum anderen solche, die sich auf die individuellen Verhaltensweisen konzentrieren. Erklärungen vom ersten, „systemischen“ Typ begegnen uns in ihrer klassischen Form als „invisible hand“ Erklärungen (vgl. zu den Originalfundstellen Schneider, 1967, und für eine grundsätzliche moderne Analyse Nozick, 1974). Diese Art der „Erklärung“ wird besonders augenfällig illustriert vom Eigen-Schelling-Modell zellulärer Automaten für die Entstehung von Ordnungsmustern in der Natur und der Gesellschaft (vgl. dazu Eigen und Winkler, 1981, Schelling, 1978, Hegselmann, 1997, Flache und Hegselmann 1998). In der experimentellen Ökonomik gehören Nachweise der Konvergenz zum Konkurrenzgleichgewicht auf experimentellen „Auktions“-Märkten in diesen Zusammenhang. Immer wieder werden in Experimenten relativ robuste Muster deutlich, nach denen sich kompetitive Systeme zu verhalten scheinen. Was in diesen geschieht, liegt gleichsam in den objektiven Situationseigenschaften begründet und ist weitgehend unabhängig von den spezifischen Annahmen über individuelles Verhalten und die subjektiven Wahrnehmungen und Bewertungen der Individuen. Der Nachweis etwa, dass sogenannte „zero intelligence traders“, die gerade nicht strategisch handeln können, auf „double auction markets“ trotzdem zum Konkurrenzgleichgewicht gelangen (vgl. Gode und Sunder, 1993), lässt den vorangehenden Punkt für eine Kombination von Simulations- und Experiment-Überlegungen besonders augenfällig werden.

In Überlegungen, die sich in der geschilderten Weise auf die Gesamteigenschaften des Interaktionssystems beziehen, kann durch die strategischen Fähigkeiten und Verhaltensweisen der Individuen gleichsam „gekürzt“ werden. Wenn man die Selektions- und Adaptationseigenschaften des Systems verstanden hat, dann weiß man, welche Muster im Prinzip entstehen können. Man kommt zu Muster-Erklärungen oder „Erklärungen im Prinzip“ (wie sie vielen ökonomischen Klassikern und im letzten Jahrhundert vor allem v. Hayek vorschwebten). Dem trägt der evolutionäre Aspekt im indirekt evolutionären Ansatz Rechnung. Es läuft danach gleichsam hinter dem Rücken der Individuen und ihrer bewusst strategischen Entscheidungen ein Selektions- und Anpassungsmechanismus ab, der ihre „subjektiv motivierten“ Entscheidungen einer „objektiven Erfolgsbewertung“ unterzieht, aber auch die individuellen Bewertungen und damit die weiteren Entscheidungen indirekt beeinflusst.

Ein ähnlicher Gedanke lag bereits Armen Alchians evolutionärer Erklärung für Verhalten, das dem eines Homo oeconomicus entspricht, zu Grunde (vgl. Alchian, 1950). Er war der Auffassung, dass er mit dem Verweis auf das Ablaufen eines Selektionsprozesses, der Opportunisten systematisch bevorzugen muss, eine Erklärung für das Auftreten opportunistischen Verhaltens auf Märkten und damit die „Erklärungsrolle“ des Homo oeconomicus gegeben habe. Sein Argument ist anregend. Es muss aber im Lichte des bislang gesagten in zwei entscheidenden Hinsichten modifiziert werden: 1. Es erklärt im Gegensatz zu den hier angestellten Überlegungen nicht systematisch das Auftreten gemischter Populationen, in denen gerade nicht das Homo oeconomicus Verhalten vorherrscht. 2. Wenn das Homo oeconomicus Verhalten keineswegs zwingend aus den evolutionären Selektionsprozessen abgeleitet werden kann, sondern in Abhängigkeit von den Informations- bzw. institutionellen Bedingungen ganz andere Verhaltensweisen das Ergebnis von Konkurrenzprozessen sein können (gerade weil sich generelle Dispositionen oder Neigungen zu solchen Verhaltensweisen lohnen), dann muss das gesamte Feld der Verhaltensweisen als möglich in Betracht gezogen und empirisch aufgeklärt werden.

Der indirekt evolutionäre Ansatz kann hier eine Integrationsleistung erbringen, die Einsichten in „objektive Konkurrenzbedingungen“ mit subjektiven Verhaltensweisen verknüpft. Inwieweit sich daraus unter Umständen bestimmte

Fingerzeige für mögliche Politiken ergeben, lässt sich an den hier diskutierten Bei-Spielen recht gut demonstrieren. Konkret gesprochen wird man die Informationskosten über den Faktor „i“ mit institutionellen Mitteln senken müssen, um entsprechende retributive Anlagen durch Systemregeln zu bewahren bzw. zu unterstützen. Wenn hinreichend viele Interaktionen so beschaffen sind, dass die Information über die Verhaltensdispositionen anderer nicht rein privater Natur ist, dann lohnt es sich, die Neigung zur Reziprozität, zum „wie Du mir, so ich Dir!“ zu kultivieren. Die Verschiedenheit sozialer Situationen zusammen mit der Verhaltensträgheit und Unfähigkeit menschlicher Individuen, beliebig genau zwischen Situationen zu diskriminieren, legt es überdies nahe, dass die betreffende Neigung nicht nur exakt in den Situationen wirksam wird, in denen sie erworben bzw. direkt verstärkt wird.

Eine Argumentation wie die vorangehende setzt voraus, dass der Erwerb von Verhaltensneigungen, die ganze Klassen von Handlungen betreffen, möglich ist und dass es Mechanismen gibt, die den Erwerb bzw. die Bewahrung dieser Neigungen in Abhängigkeit von deren positivem objektivem Wert für ihren Träger bewirken können. Wie das konkret möglich sein kann, darüber haben wir wenig gesagt, sondern nur aufgezeigt, wie die Ergebnisse solcher Kausalprozesse in einem indirekt evolutionären Modell in Zusammenhang mit der eher traditionellen entscheidungslogischen Analyse und subjektiven Situationswahrnehmung der Individuen gebracht werden können. Wenn allerdings der Erwerb von Verhaltensneigungen angenommen werden darf, dann lassen sich die klassische ökonomische Intuition, dass es häufig weit mehr auf die Systembedingungen einer Interaktion als auf die individuellen Motive ankomme (nicht der Menschenfreundlichkeit des Bäckers oder Metzgers verdanken wir unser Brot und Fleisch...), mit Modellen individuellen Verhaltens systematisch verknüpfen.

Das trägt auch den Bedürfnissen des experimentellen Ökonomen Rechnung. Denn zwei Seelen wohnen in der Brust des experimentellen Ökonomen. Zum einen versteht er als Ökonom, wie Märkte und Konkurrenz funktionieren (bzw. Institutionen wie in Young, 1998). Solche prinzipiellen Auffassungen über kompetitive Prozessen der Verhaltensselektion und –adaptation spiegeln sich auf der „oberen Ebene“ eines indirekt evolutionären Vorgehens wider. Zum anderen hat jeder experimentelle Ökonom Experimente gemacht, die ihm bei eigener

Unvoreingenommenheit unweigerlich die Folgerung aufdrängen müssten, dass das Modell unbeschränkt rationaler Verfolgung materieller egoistischer Interessen als universelles Verhaltensmodell auf der unteren Ebene der Erklärung individuellen Einzelverhaltens nichts taugt.

Entweder muss im klassischen Homo-oeconomicus-Modell die Einschränkung auf materielle eigeninteressierte Motive aufgegeben werden oder aber die Annahme „opportunistisch rationaler“ Zielverfolgung, wonach der rationale Akteur jede Chance zur Verbesserung des Zielerreichungsgrades unmittelbar wahrnimmt. Beide Strategien führen letztlich zu der Forderung nach einer „Psychologisierung der Sozialtheorie“. Allgemeine methodologische Gründe, wie sie vor allem von Hans Albert immer wieder zwingend vorgebracht wurden, stützen diese Forderung ebenso wie die Ergebnisse der experimentellen Ökonomik. Wenn die Forderung nach einer „Psychologisierung der Sozialtheorie“ dennoch von vielen Ökonomen abgelehnt wird, dann wohl deshalb, weil sie die Entwertung ihres eigenen entscheidungslogischen Humankapitals oder einen Verlust disziplinärer Eigenständigkeit befürchten.⁷

Systematisch beachtlicher als derartige psychologische Faktoren sind Befürchtungen, jene Einheitlichkeit der Theoriebildung zu verlieren, die das Pfund bildet, mit dem die Ökonomik in der Konkurrenz mit anderen Sozialwissenschaften stets wuchern wollte. Aus unserer Sicht erzeugte der reine „Rationalwahlansatz“ jedoch eher eine Einheitlichkeitsillusion auf Kosten des Realismus als eine systematische Einheitlichkeit. Die hier vorgeschlagene Modellierung mittels des indirekten evolutionären Ansatzes gibt demgegenüber an, wie man im Prinzip zu einer Vereinigungstheorie kommen könnte, die sowohl Rationalwahl- als auch adaptive Elemente kennt. Dies sollte die Verständigung unter unterschiedlich spezialisierten Sozialforschern erleichtern. Damit lässt sich der vermutlich größte Vorteil der zeitgenössischen Ökonomik, nämlich der einer halbwegs einheitlichen, formal präzisierten theoretischen Sprache, in eine Sozialtheorie hinüberretten, die vom Schatten der Vergangenheit *und* der Zukunft geprägtes menschliches Handeln erhellt. Aber

⁷ Die Möglichkeit einer Entwertung des angesammelten Methodenwissens ist dabei im Licht der Chancen zu sehen, die formale Modelle beschränkt rationaler Interaktion oder auch der hier betrachteten Evolution nicht-opportunistischer Präferenzen mathematisch-logisch orientierten Ökonomen bieten.

vielleicht sollte man auch diesen Vorzug nicht überbewerten und insgesamt unter dem Slogan „bounded rationality“ mit der kognitiv-psychologischen Fundierung der Ökonomik in einem Verhaltensmodell, das auf Modellen von realen mentalen Prozessen aufbaut, ernst machen.

Literatur

Albert, H. *Marktsoziologie Und Entscheidungslogik*. Neuwied/Berlin.: Luchterhand, 1967.

Alchian, A. A. (1950). „Uncertainty, Evolution, and Economic Theory.“ *Journal of Political Economy*, Vol. 58, pp. 211 ff.

Baurmann, M. (1996). *Der Markt Der Tugend*. Tübingen: Mohr.

Brennan, H. G. und Buchanan, J. M. (1985). *The Reason of Rules*. Cambridge: Cambridge University Press.

Cameron, L. (1999). “Raising the Stakes in the Ultimatum Game: Experimental Evidence from Indonesia.” *Economic Inquiry* 37, 47-59.

Dufwenberg, M., U. Gneezy, W. Güth und E. van Damme (2001). „Direct Versus Indirect Reciprocity: An Experiment.“ *Homo oeconomicus*, XVIII(1/2), 29-30.

Eigen, M. und Winkler, R. (1981). *Das Spiel*. München: Piper.

Fehr, E. und K.M. Schmidt (1999). „A Theory of Fairness, Competition, and Cooperation.” *Quarterly Journal of Economics* 114, 817-868.

Flache, A. und Hegselmann, R. (1998) „Understanding Complex Social Dynamics – A Plea For Cellular Automata Based Modelling.“ *Journal of Artificial Societies and Social Simulation*, 3.

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.

Frank, R. (1992). *Die Strategische Rolle Der Emotionen*. Oldenbourg: Oldenbourg Verlag.

Gode, D. K. und Sunder, S. (1993). „Allocative Efficiency of Markets with Zero Intelligence Traders: Markets as a Partial Substitute for Individual Rationality.“ *Journal of Political Economy*, 101, 119-37.

Granovetter, M. (1985). „Economic Action and Social Structure: The Problem of Embeddedness.“ *American Journal of Sociology*, 91(3), 481-510.

Güth, W. (1995). On Ultimatum Bargaining Experiments – A Personal Review. *Journal of Economic Behavior and Organization* 27, 329-344.

Güth, S., Güth, W. und Kliemt, H. (2000). „The Dynamics of Trustworthiness Among the Few.“ Veröffentlichung der Humboldt-Universität zu Berlin, Wirtschaftswissenschaftliche Fakultät. (Discussion Paper, Economics Series; 164, forthcoming in *Japanese Economic Review*)

Güth, W. und Kliemt, H. (1995 a) “On the justification of strategic equilibria - rationality requirements versus conceivable adaptive processes.” Discussion paper - Economics Series 46, Humboldt Universität zu Berlin.

Güth, W. und Kliemt, H. (1995 b). „Ist die Normalform die normale Form?“ *Homo oeconomicus* Bd. XII(1/2), 1995, 155-183.

Güth, W. und Kliemt, H. (2000). “Evolutionarily Stable Co-operative Commitments.” *Theory and Decision* 49, 197-221

Güth, W., C.Schmidt, und M. Sutter (2001) „Fairness in the Mail and Opportunism in the Internet a Newspaper Experiment on Ultimatum Bargaining,“ Berlin: Institute for Economic Theory III, Humboldt University.

Güth, W., R. Schmittberger und B. Schwarze (1982). „An Experimental Analysis of Ultimatum Bargaining.“ *Journal of Economic Behavior and Organization* 3, 367-388.

Hegselmann, R. und H. Kliemt (1997, Hrsg.). *Moral und Interesse*. München: Oldenbourg.

Hicks, J. (1979). *Causality in Economics*. Oxford: Blackwell.

Kirchsteiger, G. (1994): The role of envy in ultimatum games, *Journal of Economic Behavior and Organization* 25, 373-389.

Klein, D.B. (1997, Hrsg.) *Reputation*. Ann Arbor: The University of Michigan Press.

- Kliemt, H. (1993). „Constitutional Commitments.“ In: P. Herder Dorneich et. al. (Hrsg.), *Jahrbuch Für Neuere Politische Ökonomie*, 145-73.
- Kuhn, H.W. (1953). „Extensive Games and the Problem of Information.“ In: H.W. Kuhn und A.W. Tucker (Hrsg.), *Contributions to the Theory of Games*, Princeton, NJ: Princeton University Press, 193-216.
- Mackie, J. L. (1985). „Morality and the Retributive Emotions.“ In: ders. *Persons and Values*. Oxford: Oxford University Press.
- Nozick, R. (1974). *Anarchy, State, and Utopia*. New York: Basic Books.
- Rabin, M. (1993). „Incorporating Fairness into Game Theory and Economics.“ *American Economic Review*, 83, pp. 1281-302.
- Roth, A.E. (1995). Bargaining Experiments. In: J. Kagel und A.E. Roth (Hrsg.), *Handbook of Experimental Economics*, Princeton, NJ: Princeton University Press, 253-348.
- Schelling, T. C. (1978). *Micromotives and Macrobehavior*. New York, London: W.W. Norton & Company.
- Schneider, L. (ed.) (1967). *The Scottish Moralists on Human Nature and Society*. Chicago und London.
- Selten, R. (1965). „Spieltheoretische Behandlung eines Oligopolmodells mit Nachfrageträgheit.“ *Zeitschrift für die gesamte Staatswissenschaft* 121, 301-324 und 667-689.
- Selten, R. (1975). „Reexamination of the Perfectness Concept for Equilibrium Points in Extensive Games.“ *International Journal of Game Theory* 4, 25-55.
- Selten, R. (1983) „Evolutionary Stability in Extensive Two-Person Games.“ *Mathematical Social Sciences*, 5, pp. 269-363.
- Selten, R. (1988) „Evolutionary Stability in Extensive Two-Person Games -- Corrections and Further Development.“ *Mathematical Social Sciences*, 16(3), pp. 223-66.
- Siegart, G. (1997) „Explikation“. in W. Löffler and E. Runggaldier, *Dialog und System*. Sankt Augustin

Slonim, R. und A.E. Roth (1998). „Learning in High Stakes Ultimatum Games: An Experiment in the Slovak Republic.” *Econometrica* 66, 569- 596.

Strawson, P. F. (1962) „Freedom and Resentment." *Proceedings of the British Academy*“: 187-211.

Vanberg, V. (1981). *Liberaler Evolutionismus Oder Vertragstheoretischer Konstitutionalismus?* Tübingen.

Young, H. P. (1998). *Individual Strategy and Social Structure. An Evolutionary Theory of Institutions*. Princeton: Princeton University Press.