



Walter-Adolf-Jöhr-Vorlesung 2006

Prof. Dr. Dr. h.c. Ernst Fehr

Neuroökonomik

Die Erforschung der biologischen Grundlagen
des menschlichen Sozialverhaltens

Forschungsgemeinschaft für Nationalökonomie (Hrsg.)

Walter-Adolf-Jöhr-Vorlesung 2006

an der Universität St. Gallen

Prof. Dr. Dr. h.c. Ernst Fehr

Neuroökonomik

**Die Erforschung der biologischen Grundlagen
des menschlichen Sozialverhaltens**

St. Gallen, August 2006

Herausgeberin: Forschungsgemeinschaft für Nationalökonomie
an der Universität St. Gallen
Bodanstrasse 1, CH-9000 St. Gallen
Internet www.fgn.unisg.ch
Telefon 0041 071 224 23 11
Fax 0041 071 224 28 74

Bezug: Nicht im Buchhandel erhältlich
Bezug über die Herausgeberin

Auflage: 1'200

Copyright: Forschungsgemeinschaft für Nationalökonomie,
August 2006

Druck: Difo-Druck OHG
Laubanger 15, DE-96052 Bamberg

Walter-Adolf-Jöhr-Vorlesung an der Universität St. Gallen



Walter Adolf Jöhr

Die Walter-Adolf-Jöhr-Vorlesungen finden seit 1988 in jährlichem Abstand an der Universität St. Gallen statt. Ins Leben gerufen wurde diese Vortragsreihe zu Ehren von Professor Dr. Walter Adolf Jöhr (1910 - 1987), der von 1937 an fünfzig Jahre lang an der Universität St. Gallen wirkte. Als Rektor nahm Professor Jöhr in den Jahren 1957 bis 1963 entscheidenden Einfluss auf die Entwicklung der Universität. Die Forschungsgemeinschaft für Nationalökonomie an der HSG wurde von Professor Jöhr gegründet und massgeblich geprägt. Dem Wissenschaftler Walter Adolf Jöhr haben wir verschiedene bahnbrechende Publikationen aus dem Bereich der Volkswirtschaftslehre zu verdanken.

Prof. Dr. Dr. h.c. Ernst Fehr*

Neuroökonomik

Die Erforschung der biologischen Grundlagen des menschlichen Sozialverhaltens¹

Die Neuroökonomik vereint Methoden der Neurowissenschaften und der Ökonomie, um besser zu verstehen, wie das menschliche Gehirn in ökonomischen und sozialen Kontexten Entscheidungen trifft. Die Neuroökonomik ist Teil der allgemeinen Suche der Volkswirtschaftslehre nach Mikrofundierungen, in diesem Fall nach Mikrofundierung individuellen Entscheidungsverhaltens in sozialen Zusammenhängen. Das ökonomische Modell individueller Entscheidungsfindung basiert auf drei Konzepten: den Handlungsmöglichkeiten, den Präferenzen und den Einschätzungen über das Verhalten anderer Akteure (Beliefs). Ökonomen gehen davon aus, dass bei einer gegebenen Menge an Handlungsalternativen und gegebenen Beliefs über die Zustände der Welt und die Handlungen der anderen Spieler ein Individuum die nutzenmaximierende Handlung wählen wird.

Die Neuroökonomik ermöglicht eine Mikrofundierung von individuellen Beliefs, Präferenzen und Verhalten, indem sie die im Gehirn stattfindenden Prozesse untersucht, die mit der Entstehung von Beliefs, der Wahrnehmung der Handlungsmöglichkeiten und der individuellen Entscheidung einhergehen. Da der Handlungsspielraum unterschiedlich gestaltet sein kann und unterschiedliche Rahmenbedingungen unterschiedliches Verhalten hervorrufen können, kann die Neuroökonomik auch zu einem besseren Verständnis solcher Gestaltungseffekte beitragen.

Dieser Aufsatz diskutiert neuere neuroökonomische Erkenntnisse über altruistisches Verhalten und über das oft zu beobachtende menschliche Vertrauen in das uneigennütziges Verhalten anderer. Diese Befunde stützen die Ansicht, dass Menschen (i) aus wechselseitiger Kooperation in Sozia-

* Institut für Empirische Wirtschaftsforschung, Universität Zürich

¹ Dieser Aufsatz beruht auf mehreren Arbeiten die ich gemeinsam mit Urs Fischbacher und Michael Kosfeld durchgeführt habe, insbesondere auf Fehr, Fischbacher und Kosfeld (2005). Ich bedanke mich bei beiden dafür, dass ich das Material aus diesen Aufsätzen für diese Arbeit verwenden durfte.

len-Dilemma-Spielen (SD-Spielen) und (ii) aus der Bestrafung unfairen Verhaltens in solchen Spielen nichtpekuniären Nutzen ziehen. Somit ist wechselseitige Kooperation, welche trotz starker Anreize zum Trittbrettfahren stattfindet, und die Bestrafung von Trittbrettfahrern in SD-Spielen nicht irrational, sondern bei Vorliegen entsprechender sozialer Präferenzen eine rationale Handlung. Schliesslich präsentieren wir die Resultate einer aktuellen Studie, welche den Einfluss des Neuropeptids Oxytozin auf das Verhalten von Probanden in einem sequentiellen SD-Spiel untersucht. Im Zentrum steht hierbei die Fragestellung, auf welche Weise Oxytozin das Vertrauen zwischen den Probanden und ihre Vertrauenswürdigkeit beeinflusst. Untersuchungen in der Tierwelt haben gezeigt, dass das Hormon Oxytozin prosoziales Annäherungsverhalten fördert, was nahe legt, dass Oxytozin auch das prosoziale Sozialverhalten von Menschen beeinflussen könnte. In der Tat zeigt unsere Studie, dass Teilnehmer, denen Oxytozin verabreicht wurde, ihrem Gegenüber mehr vertrauten, obwohl Oxytozin ihre Beliefs über das Verhalten der anderen Teilnehmer nicht zu beeinflussen scheint. Dies deutet darauf hin, dass Oxytozin einen direkten Einfluss auf bestimmte Aspekte der sozialen Präferenzen der Subjekte hat. Interessanterweise hat Oxytozin, obwohl es eine Auswirkung auf das gezeigte Vertrauen hat, keinen Effekt auf die Vertrauenswürdigkeit der Subjekte.

Im allgemeinen scheut sich die ökonomische Theorie, spezifische Annahmen über Präferenzen zu treffen, mit Ausnahme der Annahme, dass Präferenzen die Axiome der Theorie der „offenbarten Präferenzen“ erfüllen. In der Praxis jedoch treffen Ökonomen häufig die starke Annahme, dass individuelle Präferenzen ausschliesslich selbstbezogen sind. Es existiert jedoch eine grosse Anzahl von Studien (Camerer, 2003, Fehr and Fischbacher, 2003), welche darauf hindeuten, dass viele Menschen soziale Präferenzen aufweisen und dass sogar noch mehr auf das Vorhandensein solcher Präferenzen vertrauen.

Sequentielle SD-Spiele, die nur einmal gespielt werden, bieten eine gute Möglichkeit, die Relevanz sozialer Präferenzen für das menschliche Verhalten zu demonstrieren. Ein sequentielles SD-Spiel kann folgendermassen beschrieben werden: Zwei Spieler, A und B, haben beide eine Anfangsausstattung von €10. Zunächst entscheidet Spieler A, ob er seine Ausstattung behalten oder sie an Spieler B transferieren möchte. Spieler B beobachtet das Verhalten von A und entscheidet danach, ob er seine (nun möglicherweise vergrösserte) Ausstattung behalten oder sie an Spieler A

senden möchte. Der Versuchsleiter verdoppelt bei jedem Transfer die transferierte Summe, d.h. beide Spieler stellen sich besser, wenn sie die ihnen verfügbare Geldsumme transferieren, als wenn sie diese für sich behalten. Diese Situation bildet einen sequentiellen ökonomischen Tausch ab, der nicht vertraglich durchgesetzt werden kann. Spieler B hat einen starken Anreiz, seine Ausstattung zu behalten, unabhängig davon ob A sein Vermögen transferiert hat oder nicht; wenn A dieses Verhalten antizipiert, besteht für ihn kein rationaler Grund seine Ausstattung zu transferieren. Ein beidseitig vorteilhafter Tausch kann somit nur stattfinden, wenn A darauf vertraut, dass sich B uneigennützig verhalten und seine Ausstattung transferieren wird und wenn B dies auch tatsächlich tut.

Hunderte von Experimenten mit Ausstattungen von bis zu mehreren Monatsinkommen bestätigen, dass ein grosser Prozentsatz der Versuchsteilnehmer in der Rolle von Spieler B Spieler A's Vertrauen erwidert, und dass ein noch grösserer Prozentsatz von Probanden in der Rolle von Spieler A Spieler B vertraut (Camerer, 2003, Fehr and Fischbacher, 2003). Wenn wir diesem Spiel nun eine dritte Entscheidungsstufe hinzufügen, indem wir Spieler A die Option geben, unter Aufwendung eigener Kosten Spieler B nachträglich zu belohnen oder zu bestrafen, belohnen viele Spieler A diejenigen Spieler B, welche ihr Vertrauen erwidern und bestrafen diejenigen Spieler B, welche ihr Vertrauen missbrauchen (Fehr, Gächter and Kirchsteiger, 1997). Weshalb beobachten wir diese starken Abweichungen von den Prognosen des konventionellen ökonomischen Modells? Was sind die treibenden Kräfte hinter der Entscheidung zu vertrauen, vertrauenswürdig zu sein und Vertrauensmissbrauch zu bestrafen? Inwiefern spielen dabei emotionale Faktoren eine Rolle und wie interagieren diese mit der menschlichen Fähigkeit, rational abzuwägen? Können Reziprozität und Bestrafungsverhalten am besten modelliert werden, indem man annimmt, sie seien Verhaltensweisen, die auf sozialen Präferenzen beruhen oder sind diese Verhaltensweisen nur eine Widerspiegelung der beschränkten Rationalität der Subjekte, wie einige Autoren meinen (Samuelson, 2005). Im Folgenden werden wir zeigen, dass neuroökonomische Studien uns helfen können, diese Fragen zu beantworten.

I. Warum bestrafen viele Menschen unfaires Verhalten?

Zusammen mit mehreren Koautoren (de Quervain et al., 2004) untersuchte ich ein sequentielles SD-Spiel mit zwei Spielern mittels der Positronenemissionstomographie (PET): PET ist eine bildgebende Methode zur Messung und Darstellung der Aktivierung verschiedener Hirnregionen. In diesem Spiel hatte Spieler A die Möglichkeit, Spieler B zu bestrafen, indem er, nachdem er beobachten konnte, ob B sein Vertrauen missbraucht hatte, an Spieler B bis zu 20 Bestrafungspunkte vergeben konnte. Die monetären Konsequenzen der Bestrafung hingen von verschiedenen Bedingungen ab, auf welche wir später näher eingehen werden. Das Gehirn von Spieler A wurde sowohl zum Zeitpunkt des Informationserhalts über B's Entscheidung, als auch während der eigenen Entscheidung, B zu bestrafen oder nicht, mittels PET beobachtet. Das Hauptziel der Studie war es, zu untersuchen, welche Prozesse in A's Gehirn ablaufen, wenn B sein Vertrauen missbraucht. Die Studie ging von der Hypothese aus, dass Spieler A ein Bedürfnis hat, B zu bestrafen, wenn B sein Vertrauen bewusst missbraucht. Diese Hypothese resultiert aus Modellen sozialer Präferenzen und Reziprozität, welche in den letzten 5-8 Jahren entwickelt wurden. Sollte diese Hypothese stimmen, erwarten wir eine Aktivierung von belohnungsbezogenen Hirnarealen während und nach Spieler A's Entscheidung, B zu bestrafen. Diese Aktivierung der Belohnungszentren könnte infolge der Befriedigung entstehen, welche ein Spieler erwartet, wenn er sich entscheidet, Spieler B für dessen unfaires Verhalten zu bestrafen.

Eine wichtige Voraussetzung für die Durchführung dieser Studie ist neurowissenschaftliches Wissen über die Hauptkomponenten neuronaler Belohnungskreisläufe im Gehirn. Glücklicherweise haben viele Studien gezeigt, dass das Striatum, ein Gebiet im Mittelhirn, ein Hauptbestandteil des neuronalen Belohnungszentrums ist. Das Erfassen einzelner Neuronen bei nicht-menschlichen Primaten (Schultz, 2000) und bildgebende Studien mit Menschen, in denen Geld als Mittel zur Belohnung eingesetzt wurde (O'Doherty et al., 2004), weisen eindeutig darauf hin, dass das Striatum ein Hauptbestandteil des neuronalen Belohnungszentrums ist. Wenn A Spieler B bestraft, weil er antizipiert, dass er durch das Bestrafen Befriedigung erlangen wird, sollte man des Weiteren vorwiegend eine Aktivierung in denjenigen Belohnungsarealen beobachten, die in Zusammenhang mit zielorientiertem Verhalten stehen. Neuronenerfassungen bei nicht-menschlichen Primaten

(Schultz, 2000) zeigen, dass das dorsale Striatum ein entscheidender Faktor für die Integration von Belohnungsinformation und Verhaltensinformation im Sinne eines zielorientierten Mechanismus ist. Weitere aktuelle bildgebende Studien unterstützen die Ansicht, dass das dorsale Striatum auch bei der Verarbeitung von Belohnungen, die aus Entscheidungen resultieren, miteinbezogen wird (O'Doherty et al., 2004). Die Tatsache, dass das dorsale Striatum auch auf erwartete monetäre Gewinne auf parametrische Weise reagiert, ist aus ökonomischer Sicht besonders interessant: wenn Probanden eine Aufgabe, welche monetäre Gewinne generiert, erfolgreich bewältigen, steigt die Aktivierung im dorsalen Striatum mit dem Anstieg des erwarteten monetären Gewinns. Wenn Spieler A's dorsales Striatum bei der Bestrafung von B aktiviert wird, haben wir demzufolge einen Hinweis darauf, dass Bestrafung unfairen Verhaltens belohnend wirkt.

Um die Aktivierung des Striatums während der Bestrafungsentscheidung zu beobachten, wurden die Gehirne der Probanden hauptsächlich in Versuchen gescannt, in denen B das Vertrauen von A missbrauchte.² In der Variante „costly“ (C) war die Bestrafung für beide Spieler A und B kostspielig. Jeder Bestrafungspunkt, den A Spieler B zuwies, kostete Spieler A €1 und reduzierte Spieler B's Auszahlung um €2. In der Variante „free“ (F), war das Bestrafen für A kostenlos, während B's Auszahlung mit jedem Bestrafungspunkt um €2 reduziert wurde. In der dritten Variante, welche wir mit „symbolic“ (S) bezeichnen, hatte die Bestrafung nur einen symbolischen Wert und keine monetären Konsequenzen. Jeder Bestrafungspunkt, der B zugewiesen wurde, verursachte weder für A noch für B monetäre Kosten. Somit konnte Spieler A Spieler B's Auszahlung in dieser Variante nicht reduzieren.

Die Hypothese, dass Bestrafung belohnend wirkt, lässt vermuten, dass der Unterschied im PET-Scanbild zwischen der F- und S-Variante die relative Aktivierung von belohnungsbezogenen Hirnregionen zeigen wird, nachdem A's Vertrauen missbraucht wurde.³ Die Überlegung hinter dieser Prognose ist, dass A zwar wahrscheinlich sowohl in der F- als auch in der S-Variante das Bedürfnis hat, B zu bestrafen, nachdem sein Vertrauen von B missbraucht wurde, er aber in der S-Variante Spieler B keinen wirklichen Schaden zufügen kann. Demzufolge ist die rein symbolische Bestrafung in der

² Spieler A spielte das Spiel sieben Mal mit sieben verschiedenen Subjekten in der Rolle von Spieler B.

³ Hirn-Aktivierungen in bildgebenden Studien werden immer relativ zu der anderen Variante gemessen. Demzufolge enthält der F-S-Unterschied die Information über jene Hirnregionen, welche in der F-Variante stärker aktiviert werden als in der S-Variante.

S-Variante kaum befriedigend, da das Bedürfnis, den Defektor zu bestrafen, nicht wirkungsvoll befriedigt werden kann. Im unwahrscheinlichen Fall, dass eine symbolische Bestrafung Befriedigung bringt, kann davon ausgegangen werden, dass sie geringer ist als in der F-Variante.

Der F-S-Unterschied eignet sich gut für die Untersuchung der aus wirkungsvoller Bestrafung resultierenden Befriedigung, denn die F- und die S-Variante unterscheiden sich ja nur in der Wirksamkeit der verhängten Bestrafung. Es ist jedoch so, dass Bestrafung aus ökonomischer Sicht auch dann befriedigend sein sollte, wenn sie etwas kostet. Wenn wirklich ein Bedürfnis nach Bestrafung von Defektoren vorhanden ist und die Versuchsteilnehmer auch tatsächlich bestrafen, ist der Akt der Bestrafung wie der Kauf eines Gutes zu interpretieren. Rationale Akteure erwerben zusätzliche Einheiten eines Gutes solange die marginalen Kosten der zusätzlichen Einheit niedriger sind als ihr marginaler Nutzen. Somit prognostiziert ein ökonomisches Modell, welches von einem Bedürfnis nach Bestrafung ausgeht, dass Bestrafung in der kostenverursachenden C-Variante auch als befriedigend empfunden werden sollte, solange die Kosten der Bestrafung nicht zu hoch sind und impliziert damit, dass belohnungsbezogene Regionen auch in der C-Variante aktiviert werden sollten.

Auswertungen von Fragebögen und das Verhalten der Probanden weisen darauf hin, dass Spieler A tatsächlich ein starkes Bedürfnis hat, Defektoren zu bestrafen. Fast alle Probanden wählen in der F-Variante die maximal mögliche Bestrafung und die meisten bestrafen auch noch in der C-Variante, wenn auch auf einem tieferen Niveau. Diese Senkung des Bestrafungsniveaus lässt sich dadurch erklären, dass Bestrafung in der C-Variante Kosten verursacht. Wichtiger jedoch ist die Tatsache, dass das dorsale Striatum im F-S-Unterschied und im C-S-Unterschied stark aktiviert wurde, eine Beobachtung, die dafür spricht, dass der Akt der Bestrafen als befriedigend empfunden wird. Ferner zeigen die Daten, dass diejenigen Probanden, die eine stärkere Aktivierung im dorsalen Striatum aufweisen, in der C-Variante stärker bestrafen. Diese positive Korrelation kann auf zwei Arten interpretiert werden. Zum einen könnte das höhere Niveau der Bestrafung eine höhere Aktivierung des dorsalen Striatums, d.h. ein höheres Niveau der Befriedigung hervorrufen. Zum anderen könnte aber auch die durch das Bestrafen auszulösende höhere antizipierte Befriedigung das höhere Bestrafungsniveau verursachen, d.h. die Aktivierung im Striatum reflektiert die antizipierte Befriedigung des Bestrafungsaktes. Aus ökonomo-

mischer Sicht wäre es beruhigend, wenn die zweite Interpretation die Richtige wäre, denn sie basiert auf der Idee, dass die antizipierte Belohnung, die durch den Akt des Bestrafens entsteht, die Bestrafungsentscheidung beeinflusst.

Sowohl die populärwissenschaftliche Presse als auch Psychologen und Neurowissenschaftler behaupten häufig, dass Emotionen eine übermächtige Kraft seien, die rationales Verhalten blockiert. Von Emotionen, wie zum Beispiel Wut, ist bekannt, dass sie eine wichtige Rolle bei der Bestrafung von Defektoren spielen (Fehr and Gächter, 2002). Es ist also möglich, dass Wut Rationalität ausser Kraft setzt und die Probanden dazu bringt, die Defektoren „blind“ zu bestrafen. Wenn jedoch gezeigt werden könnte, dass die Probanden die Bestrafungsentscheidung trotz ihrer Wut rational treffen, könnte man zugunsten eines ökonomischen Ansatzes argumentieren. Gemäss diesem Ansatz beeinflussen Emotionen wie Wut die hedonischen Konsequenzen unterschiedlicher Handlungen; die Probanden entscheiden aber rational, indem sie die Kosten und Nutzen der Handlungen abwägen.

DeQuervain et al. (2004) liefern zwei Indizien zugunsten eines ökonomischen Ansatzes. Das erste Beweisstück bezieht sich auf den C-F-Unterschied. Die Probanden in der C-Variante unterliegen einem nicht-trivialen Zielkonflikt zwischen Nutzen und Kosten des Bestrafens; die Entscheidung in der F-Variante ist hingegen viel einfacher, da dort das Bestrafen keine Kosten verursacht. Demzufolge sollten gewisse Bereiche des präfrontalen Cortex (Brodmann Areale 10 und 11), welche dafür bekannt sind, bei der Abwägung von Nutzen und Kosten in der Entscheidungsfindung involviert zu sein, in der C-Variante stärker aktiviert werden als in der F-Variante. Dies ist in der Tat der Fall. Das zweite Beweisstück basiert auf der Beobachtung, dass die meisten Probanden in der F-Variante die maximale Strafe wählten. Somit können die individuellen Differenzen in der Striatumaktivierung nicht durch unterschiedliche Bestrafungsniveaus ausgelöst worden sein. Wenn hingegen unterschiedliche Striatumaktivierungen Unterschiede in der antizipierten Befriedigung aus dem Bestrafen reflektieren, sollten diejenigen Probanden, die in der F-Variante eine höhere Striatumaktivierungen aufweisen obwohl sie dasselbe maximale Bestrafungsniveau wählen, bereit sein, in der C-Variante mehr Geld für die Bestrafung aufzuwenden. Auch diese Prognose wird von den Daten bestätigt.

II. Schafft Kooperation nicht-pekuniären Nutzen?

Modelle über soziale Präferenzen und Reziprozität basieren auf der Idee, dass ein beachtlicher Anteil von Personen in SD-Situationen wechselseitige Kooperation gegenüber einer einseitigen Defektion bevorzugen. Diese Modelle basieren auf Verhaltensbefunden, die zeigen, dass viele B-Spieler in einem sequentiellen SD-Spiel das Vertrauen von Spieler A mit Vertrauenswürdigkeit erwidern. Dennoch argumentieren Skeptiker (Samuelson, 2005), dass dieses Verhalten, welches mit sozialen Präferenzen grundsätzlich konsistent ist, mit Eigeninteresse erklärt werden kann, wenn Probanden auch in einmaligen Interaktionsspielen quasi aus Gewohnheit Verhaltensroutinen anwenden, welche eigentlich nur in wiederholten Interaktionen rational sind.

Neuroökonomische Evidenz könnte helfen, auch diese Debatte zu klären. Eine Möglichkeit besteht darin, aufzuzeigen, dass beidseitige Kooperation einen höheren Nutzen hervorbringt als einseitige Defektion. Allerdings ist die Messung der Unterschiede in den Hirnprozessen zwischen wechselseitiger Kooperation und unilateraler Defektion nicht ideal, da Unterschiede in der Hirnaktivität auf die Tatsache zurückzuführen sein könnten, dass die gescannten Spieler in einer Situation kooperieren und in der anderen betrügen. Die gemessenen Aktivierungen haben möglicherweise nichts mit den emotionalen Konsequenzen der wechselseitigen Kooperation zu tun, sondern könnten nur aufgrund unterschiedlichen Verhaltens entstehen. Es gibt aber einen anderen Weg, um das Problem zu lösen. Eine der ersten neuroökonomischen Studien (Rilling et al., 2002) berichtet von Aktivierungen im Striatum, wenn die Spieler gegenseitige Kooperation mit einem Menschen statt mit einem Computer als Partner erfahren. Obwohl also die monetäre Auszahlung der Subjekte in beiden Situationen gleich ist, wird beidseitige Kooperation mit einem Menschen offensichtlich als belohnender empfunden; eine Beobachtung, die darauf hinweist, dass durch beidseitige Kooperation ein zusätzlicher Nutzen erlangt wird, der über den reinen monetären Gewinn hinausgeht. Leider basiert die Studie von Rilling et al. jedoch auf einem wiederholten SD-Spiel. Ein wiederholtes Dilemma-Spiel beinhaltet eine Reihe anderer, nicht abgrenzbarer Einflüsse, welche die Interpretation der Hirnaktivität im Hinblick auf soziale Präferenzen erschweren.

Ein neueres Papier, welches auf einem einmalig gespielten sequentiellen SD basiert, löst dieses Problem (Rilling et al., 2004). Die Autoren zeigen wiederum, dass das Ergebnis gegenseitiger Kooperation mit einem menschlichen Partner eine höhere Aktivierung des Striatum generiert als das Ergebnis gegenseitiger Kooperation mit einem Computer.⁴ Darüber hinaus generiert das Ergebnis beidseitiger Kooperation mit einem Menschen auch eine höhere Aktivierung als derselbe monetäre Verdienst in einer trivialen Aufgabe mit individueller Entscheidungsfindung. Eine weitere Studie zeigt, dass das bloße Betrachten der Gesichter derjenigen Spielpartner, die in einem vorangegangenen SD kooperiert hatten, die Belohnungsareale aktiviert (Singer et al., 2004), und weist somit auf die besonderen hedonischen Qualitäten gegenseitiger Kooperation hin. Dieses Resultat legt nahe, dass Menschen nicht nur einen grösseren Nutzen aus Interaktionen mit kooperativen Personen haben, weil sie in solchen Interaktionen eine höhere Auszahlung erhalten können, sondern weil diese Interaktionen per se belohnend sind.

III. Die Neurobiologie des Vertrauens

Neuroökonomik ist nicht auf die Verwendung von bildgebenden Verfahren beschränkt. Eine kürzlich veröffentlichte Studie (Kosfeld et al., 2005) untersucht die neurobiologische Grundlage von Vertrauen und Vertrauenswürdigkeit in einem sequentiellen SD-Spiel. Untersuchungen mit Tieren über die Neurobiologie gewisser Formen sozialen Verhaltens (Insel and Young, 2001) legen nahe, dass das Neuropeptid Oxytozin eine biologische Grundlage für das Vertrauen unter Menschen liefern könnte. Oxytozin erleichtert die mütterliche Fürsorge und Paarbindung bei unterschiedlichen Spezies. Insbesondere scheint Oxytozin Tieren zu ermöglichen, ihre angeborene Berührungsangst zu überwinden und ihr defensives Verhalten zu unterdrücken, wodurch Kontaktaufnahme und gemeinsame elterliche Fürsorge erleichtert wird.

Kosfeld et al. (2005) untersuchten die Hypothese, dass Oxytozin Vertrauen und Vertrauenswürdigkeit fördert, indem sie das Verhalten einer Gruppe

⁴ In den Studien von Rilling et al. (2002, 2004) wird das ventrale und nicht das dorsale Striatum aktiviert. Ein plausibles Ergebnis, da die Hirnkontraste gemessen wurden nachdem die Subjekte, welche kooperiert hatten, darüber informiert wurden, ob ihr (menschliches oder computer-gesteuertes) Gegenüber auch kooperiert hat. Somit misst der PET-Unterschied den erfahrenen und nicht den antizipierten zusätzlichen Nutzen aus beidseitiger Kooperation mit einem menschlichen Partner.

von Testpersonen, welchen Oxytozin verabreicht wurde, mit dem Verhalten einer Kontrollgruppe verglichen, die ein Placebo erhalten hatten. Die Resultate zeigen, dass die Oxytozin-Gruppe ein signifikant höheres Vertrauen aufweist; auf Spieler B's Vertrauenswürdigkeit hingegen hatte Oxytozin keinen Einfluss. Konkret stieg unter dem Einfluss von Oxytozin der Prozentsatz der Spieler A, die maximales Vertrauen in einem SD gezeigt hatten, von 21 Prozent auf 45 Prozent, während die Transfers von Spieler B konstant blieben. Kosfeld et al. messen ebenfalls, wie Oxytozin die Ruhe und die Stimmung der Probanden beeinflusst, um zu kontrollieren, ob solche Nebenwirkungen möglicherweise für den Effekt von Oxytozin auf das Vertrauen verantwortlich sind. Auch bei Kontrolle für diese indirekten Effekte bleibt ein beträchtlicher und signifikanter Effekt von Oxytozin auf das Vertrauen bestehen. Der direkte Effekt von Oxytozin erhöht die Wahrscheinlichkeit, dass maximal vertraut wird, um 20 Prozentpunkte.

Eine interessante Frage ist, ob Oxytozin auf der Ebene der Beliefs der Probanden über die Vertrauenswürdigkeit der anderen Spieler wirkt oder auf der Ebene der Präferenzen. Eine aktuelle Studie (Bohnet and Zeckhauser, 2004) zeigt, dass die Entscheidung, zu vertrauen, nicht nur durch Risikoaversion beeinflusst wird, sondern auch durch die Aversion davor, ausgenutzt zu werden, d.h. durch die Angst, von Spieler B getäuscht zu werden. Demzufolge könnte Oxytozin, ebenso wie es die angeborene Tendenz von Tieren, andere Tiere zu meiden, überwindet, auch die „angeborene“ Angst, in einem SD-Spiel ausgenutzt zu werden, überwinden. Die Resultate der Studie von Kosfeld et al. (2005) zeigen, dass, obwohl Oxytozin die Subjekte leicht optimistisch stimmt, Oxytozin keinen signifikanten Effekt auf die Beliefs über Spieler B's Vertrauenswürdigkeit hat. Obwohl die Oxytozin-Gruppe und die Placebo-Gruppe ähnliche Beliefs haben, setzen sich die Probanden der Oxytozin-Gruppe viel stärker einer möglichen Ausnützung aus, indem sie Spieler B mehr Geld senden. Somit scheint es möglich, dass Menschen unter Einfluss von Oxytozin eher bereit sind, das Risiko auf sich zu nehmen, ausgenutzt zu werden. Diese Beobachtung impliziert, dass Oxytozin die Aversion, ausgenutzt zu werden, beeinflusst. Dieser Effekt ist insofern interessant, als Ökonomen meistens annehmen, dass Präferenzen stabil sind. Wenn Präferenzen jedoch auf wirklichen oder antizipierten Emotionen, die starken Schwankungen unterliegen, basieren, sind Präferenzen möglicherweise viel weniger stabil als typischerweise angenommen wird. Ferner können, wie die Oxytozin-Studie suggeriert, Präferenzen und die dahinterliegenden Gemütszustände über kurze Perioden

hinweg absichtlich durch Verabreichung der „richtigen“ Substanz manipuliert werden.

IV. Schlussfolgerungen

In dieser Arbeit wurde jüngere neuroökonomische Evidenz zu sozialen Präferenzen und Vertrauen diskutiert. Die Implikationen von neuroökonomischen Studien gehen jedoch weit über diese Forschungsgebiete hinaus (Camerer, Loewenstein and Prelec, 2005). Neuroökonomische Studien ermöglichen sehr wahrscheinlich einen Einblick in die Frage, wie das Belohnungszentrum mit der Entscheidungsfindung bei intertemporalen Entscheidungen (McClure et al., 2004) und Risiko (Breiter et al., 2001) verknüpft ist, und wie Affekt und Kognition interagieren, um Entscheidungen zu generieren (Sanfey et al., 2003). Solche Studien ermöglichen es uns, über den bisherigen „als-ob“-Ansatz der Ökonomie hinauszugehen, indem sie die neuronalen Mechanismen, die hinter individuellen Entscheidungen stehen, aufdecken. Langfristig könnten neuroökonomische Einsichten den gängigen „Präferenzen-und-Beliefs“-Ansatz, der in der Ökonomie herrscht, fundamental verändern.

Ökonomen nehmen beispielsweise an, dass der Belief eines Individuums über die Handlung anderer Spieler unabhängig ist von den Präferenzen des Individuums. Diese Annahme schliesst motivierte Beliefbildung aus und macht es daher schwierig, Fragen über religiösen Glauben, Ideologien, Aggression gegenüber Mitglieder anderer Gruppen oder die Struktur und den Inhalt von politischen und ökonomischen Werbekampagnen, welche die Emotionen der Menschen ansprechen, zu verstehen. Vielleicht gibt es aber neuronale und affektive Mechanismen, welche es den Präferenzen ermöglichen, die Beliefs zu beeinflussen, und umgekehrt. Die Reputationsbildung könnte ein Beispiel liefern: wenn wir in einem sozialen Tausch betrogen werden, haben wir eine starke affektive Reaktion, welche unsere Präferenzen gegenüber dem Kontrahenten bestimmt (Singer et al., 2004). Diese affektive Reaktion könnte auch unsere Beliefs über das zukünftige Verhalten des Kontrahenten formen. Wir wären erstaunt, wenn eine solche affekt-geleitete Beliefbildung den Regeln Bayesianischen Aufdatierens von Wahrscheinlichkeiten folgen würde.

Literatur

- Bohnet, I. und Zeckhauser, R., (2004). "Trust, risk and betrayal". *Journal of Economic Behavior & Organization* 55, 467-484.
- Breiter, H. C., Aharon, I., Kahneman, D., Dale, A. und Shizgal, P., (2001). "Functional imaging of neural responses to expectancy and experience of monetary gains and losses". *Neuron* 30, 619-39.
- Camerer, C., Loewenstein, G. und Prelec, D., (2005). "Neuroeconomics: How Neuroscience Can Inform Economics". *Journal of Economic Literature* 43 1, 9-64.
- de Quervain, D. J. F., Fischbacher, U., Treyer, V., Schelthammer, M., Schnyder, U., Buck, A. und Fehr, E., (2004). "The neural basis of altruistic punishment". *Science* 305, 1254-1258.
- Fehr, E. und Fischbacher, U., (2003). "The nature of human altruism". *Nature* 425, 785-791.
- Fehr, E., Fischbacher, U., und Kosfeld M., (2005). "Neuroeconomic foundation of trust and social preferences". CEPR Discussion Papers 5127, Centre for Economic Policy Research.
- Fehr, E. und Gächter, S., (2002). "Altruistic punishment in humans". *Nature* 415, 137- 140.
- Fehr, E., Gächter, S. und Kirchsteiger, G., (1997). "Reciprocity as a contract enforcement device: Experimental evidence". *Econometrica* 65, 833-860.
- Insel, T. R. und Young, L. J., (2001). "The neurobiology of attachment". *Nat Rev Neurosci* 2, 129-36.
- Kosfeld, M., Heinrichs, M., Zak, P. J., Fischbacher, U. und Fehr, E., (2005). "Oxytocin increases trust in humans". *Nature* 435, 673-676.

-
- McClure, S. M., Laibson, D. I., Loewenstein, G. und Cohen, J. D., (2004). "Separate neural systems value immediate and delayed monetary rewards". *Science* 306, 503-7.
 - O'Doherty, J., Dayan, P., Schultz, J., Deichmann, R., Friston, K. und Dolan, R. J., (2004). "Dissociable roles of ventral and dorsal striatum in instrumental conditioning". *Science* 304, 452-454.
 - Rilling, J. K., Gutman, D. A., Zeh, T. R., Pagnoni, G., Berns, G. S. und Kilts, C. D., (2002). "A neural basis for social cooperation". *Neuron* 35, 395-405.
 - Rilling, J. K., Sanfey, A. G., Aronson, J. A., Nystrom, L. E. und Cohen, J. D., (2004). "Opposing BOLD responses to reciprocated and unreciprocated altruism in putative reward pathways". *Neuroreport* 15, 2539-2243.
 - Samuelson, L., (2005). "Foundations of Human Sociality: A Review Essay". *Journal of Economic Literature* 43 2, 488-97.
 - Sanfey, A. G., Rilling, J. K., Aronson, J. A., Nystrom, L. E. und Cohen, J. D., (2003). "The neural basis of economic decision-making in the ultimatum game". *Science* 300, 1755-1758.
 - Schultz, W., (2000). "Multiple reward signals in the brain". *Nature Reviews Neuroscience* 1, 199-207.
 - Singer, T., Kiebel, S.J., Winston, J. S., Kaube, H., Dolan, R. J. und Frith, C. D., (2004). "Brain responses to the acquired moral status of faces". *Neuron* 41, 653-62.

Walter-Adolf-Jöhr-Vorlesungen

- 1988 Prof. Dr. **Wilhelm Krelle**, Bonn
Lange Wellen der wirtschaftlichen Entwicklung
- 1989 Prof. Dr. **Ota Šik**, St. Gallen
Der Weg 3a und 3b. Neue Aussichten für eine Konvergenz der Systeme
- 1990 Prof. Dr. **Kurt W. Rothschild**, Wien
Deregulierung - Anatomie eines Schlagwortes
- 1991 Prof. Dr. **Gottfried Bombach**, Basel
Zur Arbeitslosigkeit der Gegenwart
- 1992 Prof. Dr. **Ernst Heuss**, Erlangen-Nürnberg
Evolutorik und Marktwirtschaft
- 1993 Prof. Dr. **Erich Streissler**, Wien
Unmöglichkeit und Notwendigkeit der Wirtschaftspolitik als bleibendes Dilemma
- 1994 Prof. Dr. **Wolfgang Stolper**, Ann Arbor, USA
Geschichte und Theorie in der Analyse des kapitalistischen Prozesses - ein Schumpeterscher Ansatz
- 1995 Prof. Dr. **Hans Albert**, Heidelberg
Die Idee rationaler Praxis und die ökonomische Tradition
- 1996 Prof. Dr. Dres. h.c. **Norbert Klöten**, Tübingen
Im Sog der Währungsunion. Reflektionen über die Europäische Union und die Schweiz
- 1997 Prof. Dr. **Hans Christoph Binswanger**, St. Gallen
Chinesische Ökonomik. Fünf ordnungspolitische Denkrichtungen in der Chinesischen Ökonomik und ihre europäischen Parallelen

- 1998 Prof. Dr. Dr. h.c. **Heinz König**, Mannheim
Gründungsdynamik, Innovation und Beschäftigung
- 1999 Prof. Dr. Dres. h.c. **Herbert Giersch**, Kiel
Marktökonomik für die offene Gesellschaft
- 2000 Prof. Dr. **Peter Bernholz**, Basel
Globalisierung und Umstrukturierung: Sind sie neu?
- 2001 Prof. Dr. Dres. h.c. mult. **Bruno S. Frey**, Zürich
Glück und Nationalökonomie
- 2002 Prof. Dr. **Ulrich Witt**, Jena
Wirtschaftswachstum - was geschieht auf der Nachfrageseite?
- 2003 Prof. Dr. Dr. h.c. **Horst Siebert**, Kiel
Müssen Arbeitslose sein? Zur Reform von Arbeitsmarkt und Sozialstaat in Deutschland
- 2004 Prof. Dr. Dr. h.c. **Hans-Werner Sinn**, München
Das Dilemma der Globalisierung
- 2005 Prof. Dr. **Ernst Baltensperger**, Bern
*Vom Sinn und Unsinn des Schuldenmachens
Privatschulden, Staatsschulden, Aussenschulden -
Plädoyer für langfristige Konsistenz*
- 2006 Prof. Dr. Dr. h.c. **Ernst Fehr**, Zürich
*Neuroökonomik - Die Erforschung der biologischen
Grundlagen des menschlichen Sozialverhaltens*

Texte unter

<http://www.fgn.unisg.ch/public/public.htm>