

Neuroökonomik, Institutionen und verteilte Kognition: Empirische Grundlagen eines nicht-reduktionistischen naturalistischen Forschungsprogramms in den Wirtschaftswissenschaften

von *Carsten Herrmann-Pillath*
Frankfurt School of Finance & Management

vorgelegt anlässlich der Sitzung des Ausschusses für Wirtschaftssysteme und
Institutionenökonomik beim Verein für Socialpolitik, Münster, 25.-27.9. 2011

A. Die Wahlverwandtschaft von Neuroökonomik und Institutionenökonomik aus der Sicht des Externalismus

In den letzten zehn Jahren gehört die Neuroökonomik sicherlich zu den dynamischsten Forschungsfeldern der Wirtschaftswissenschaften. Jedoch ist sie auch umstritten (*Harrison* 2008), weil es verschiedene methodische Schwachstellen gibt, vor allem aber, weil in Frage gestellt wird, dass sie einen Erkenntnisfortschritt liefern kann, was die Behandlung spezifisch wirtschaftswissenschaftlicher Fragestellungen mit ökonomischen Methoden angeht (*Gul* und *Pesendorfer* 2008). Mit der jüngsten paradigmatischen Formulierung der Neuroökonomik durch *Glimcher* (2011) hat sich dieses Problem im Grunde zugespitzt. *Glimcher* vertritt das Programm eines neuroökonomischen Reduktionismus: Er setzt sich zum Ziel, das herkömmliche Konzept des subjektiven Nutzens durch ein naturalistisches Nutzenkonzept zu ersetzen, in dem subjektive Nutzenzuschreibungen neuronalen Zuständen des Gehirns entsprechen - mit dem sicherlich wesentlichen Zusatz, als ökonomische Nutzenfunktion eine verhaltensökonomisch modifizierte zu verwenden, nämlich mit Referenz-Punkt-Effekten. Dieses Programm kann sich aber letztendlich selbst arbeitslos machen: Denn wenn eine Reduktion tatsächlich möglich ist, gibt es aus ökonomischer Sicht keinen wirklichen Anlaß mehr, neuroökonomisch erweiterte Modelle zu verwenden, denn die Konsequenzen auf der Ebene ökonomisch relevanter und erfaßbarer Phänomene sind nicht mehr von ökonomischen Standardmodellen verschieden, erneut unter Berücksichtigung verhaltensökonomischer Modifikationen (*Bernheim* 2009). Zwar würde die Neuroökonomik erklären, warum bestimmte Verhaltensweisen ökonomischer Subjekte generiert werden, aber diese Frage wäre für den ‚normalen‘ Ökonomen ähnlich irrelevant wie die Quantenmechanik für die Arbeit der meisten Praktiker in der Angewandten Chemie. Mit anderen Worten: Gerade der Erfolg des von *Glimcher* vorgeschlagenen Reduktionsprogramms würde die Wirtschaftswissenschaft letztendlich unberührt lassen, was neurowissenschaftliche Erkenntnisse angeht.

In der Tat stellt sich die Frage, inwieweit der Begriff der ‚NeuroÖKONOMIK‘ zielführend ist. *Glimcher* nimmt ihn offensichtlich ernst, indem er sich strikt am bestehenden ökonomischen Theoriegebäude orientiert. Daher vertritt er als Neurowissenschaftler auch die umgekehrte Bewegung, nämlich die Anwendung ökonomischer Modelle auf neuronale Phänomene (*Glimcher* 2003; vgl. *Ross* 2008). In vielen sogenannten ‚neuroökonomischen‘ Ansätzen geht es aber um etwas anderes: Nämlich die Nutzung neurowissenschaftlicher Erkenntnisse zur Erklärung ökonomischer Phänomene, ohne einen spezifischen Anspruch bezüglich der theoretischen Integration der beiden Disziplinen zu verfolgen. Beispielsweise wird dann auch konsequent von ‚Consumer NeuroSCIENCE‘ (*Hubert* und *Kenning* 2008; vgl. *Hubert* 2010) oder ‚Organizational NeuroSCIENCE‘ (*Becker*, *Cropanzano* und *Sanfey* 2011) gesprochen, nicht aber von speziellen Anwendungen der NeuroÖKONOMIK. Derart offene

Ansätze finden sich eher in den Managementwissenschaften, die seit jeher multidisziplinär angelegt sind (etwa der Sammelband *Stanton et al. 2010*). In den Wirtschaftswissenschaften im engen Sinne (als ‚Economics‘) findet sich diese multidisziplinäre Sicht in der Verschränkung zwischen Verhaltensökonomik bzw. experimenteller Ökonomik und Neurowissenschaften, die inzwischen besonders auch in Deutschland floriert (programmatisch umrissen von *Camerer et al. 2005*). Auch hier tritt aber das Paradox des Reduktionismus auf: Welchen zusätzlichen Erkenntnisbeitrag leistet die Neurowissenschaft über die weitergehende Erklärung bestimmter Phänomene hinaus, die bislang durch rein psychologische Konstrukte bzw. durch bestimmte Erweiterungen der ökonomischen Modelle geleistet wurde (*Camerer 2007*)? Wenn zum Beispiel bestimmte Hypothesen über ‚Soziale Präferenzen‘ entwickelt werden, die experimentell getestet werden: Welche zusätzliche Einsicht wird dadurch gewonnen, dass gezeigt wird, welche Bereiche des Gehirns aktiviert werden, wenn Handlungen unter sozialen Präferenzen stattfinden (*Fehr 2009*)?

Diese Fragen werden in den Wirtschaftswissenschaften oft eher intuitiv diskutiert, entsprechen aber letztendlich altvertrauten Themen in der analytischen Philosophie des Geistes (also der heutzutage angloamerikanisch dominierten ‚Philosophy of Mind‘): Wieweit lassen sich mentale Phänomene vollständig auf neuronale Phänomene reduzieren? Wenn ja, welche konkrete Form hat diese Reduktion? Im Zusammenhang dieser Debatten spielt die grundsätzliche Frage eine zentrale Rolle, die auch meine Diskussion der Beziehung zwischen Neuroökonomik und Institutionen orientieren soll, nämlich die Entscheidung zwischen Internalismus und Externalismus als grundsätzlichen Positionen (Überblick bei *Schanz 2004*; *Wilson 2004*; *Lau und Deutsch 2008*). Grob gesagt, vertritt der Internalist die Auffassung, dass es irreduzible mentale Phänomene gibt, wie etwa die ‚Qualia‘, die wesentlich für die Erklärung menschlichen Verhaltens sind, und die ausschließlich intern generiert werden. Der Externalist behauptet hingegen, dass menschliches Verhalten vollständig durch kausale Beziehungen zwischen Außenwelt und Gehirn erklärt werden kann.

Bei dieser einfachen Entgegensetzung wird sofort klar, dass die Neuroökonomik buchstäblich quer dazu liegt. In der Wirtschaftswissenschaft ist der subjektive Nutzen eindeutig ein ‚Qualium‘, denn er ist nicht nur irreduzibel, sondern kann auch nicht direkt gemessen oder beobachtet werden, aber genau deshalb wurde er seit der marginalistisch-subjektivistischen Revolution auch nicht mehr als Gegenstand ökonomischer Forschung betrachtet. Das empirisch relevante Konzept ist dasjenige der ‚offenbarten Präferenzen‘, die ausschließlich auf Daten über beobachtbare Wahlhandlungen beruhen (vgl. *Ross 2005: 104ff.*). Diese Wahlhandlungen werden rein behavioristisch mit einem ‚black box‘ Modell erklärt: Es gibt externe Reize, wie ein Angebot unterschiedlicher Früchte zum Verkauf, und einen Output, die Wahlhandlung. Die ‚black box‘ dazwischen wird durch die ökonomische Nutzenfunktion abgebildet, die aber keine deskriptive Rolle spielt, also keineswegs beansprucht, einen Einblick in die ‚black box‘ zu bieten: Sie ist ausschließlich als mathematische Beschreibung der beobachtbaren Wahlhandlungen intendiert. Nun: ganz offensichtlich ist der Begriff des subjektiven Nutzens internalistisch, während aber der Begriff der offenbarten Präferenzen externalistisch ist. Das bedeutet, das ökonomische Standardmodell führt zwei diametral entgegengesetzte Positionen in der Philosophie des Geistes zusammen - was natürlich eigentlich nicht haltbar ist, aber nie virulent wurde, weil die meisten Ökonomen faktisch einer instrumentalistischen Methodologie folgen, wie sie *Friedman (1953)* klassisch formuliert hatte: Der Realitätsgehalt der Grundannahmen ist für die Theorieevaluation nicht relevant, solange empirische Tests von Prognosen strikt durchgeführt werden. Wie wir nun gesehen haben, geht es gar nicht nur um den Realitätsgehalt als solchen, sondern um viel grundlegendere Positionen in der Philosophie des Geistes.

Warum ist diese Problematik wichtig, wenn es um das Verhältnis zwischen Neuroökonomik und Institutionen gehen soll? Es gibt zwei Auswege aus dem

philosophischen Dilemma des ökonomischen Standardmodells, nämlich entweder ein konsequent internalistischer oder ein konsequent externalistischer. *Glimchers* Ansatz ist konsequent internalistisch, aber gleichzeitig anti-mentalistisch, weil er einen vollständigen neuronalen Reduktionismus anstrebt. Eine solche Position ist schwierig, weil er Kausalketten verkürzt bzw. implizit eine kausale Priorität der neuronalen Prozesse unterstellt, die eigentlich nicht haltbar ist. Rein neuronale Modelle menschlichen Verhaltens sind gewöhnlich externalistisch, weil das Gehirn auf externe Reize reagiert. Internalistische Ansätze sind verfügbar und würden sich auf die Autonomie des Gehirns in seiner Beziehung zur Außenwelt stützen: Denn nur solche Gehirnaktivitäten würden eine internalistische Position stützen, die eindeutig und ausschließlich intern generiert werden (was natürlich für den Tatbestand des spontanen ‚Feuerns‘ von Neuronen gilt). *Glimcher* ignoriert diese Ansätze aber vollkommen, weil sie der mechanistischen Ontologie des ökonomischen Nutzenmodells zuwiderlaufen, das im Grunde ein behavioristisches Reiz-Reaktions-Schema ist. Solche Ansätze sind durchaus verfügbar, und interessanterweise gehört hierzu die frühe Arbeit eines Ökonomen, nämlich *Hayek* (1952). Bereits vor längerer Zeit haben die Gehirnwissenschaftler *Varela* und *Maturana* entsprechende Konzeptionen zum Gehirn als ‚autopoietisches System‘ entworfen, die auch von der Gehirnwissenschaft rezipiert wurden (z.B. *Roth* 1992). Insofern scheint *Glimchers* Ansatz weiter an der Spannung zwischen Internalismus und Externalismus zu leiden, die dem ökonomischen Referenzmodell inhärent ist.

Die Alternative ist ein rein externalistischer Ansatz. Er wäre philosophisch konsequent im Sinne, dass die Kausalketten zwischen der Umwelt und dem neuronalen System als analytische Einheit betrachtet würden (etwa im Sinne der sogenannten Teleosemantik, siehe *Millikan* 1989; *MacDonald* und *Papineau* 2006, *Neander* 2009). Daraus würde aber folgen, dass jedwede Bezugnahme auf subjektiven Nutzen obsolet würde. Stattdessen muß gefragt werden, wie bestimmte Wechselwirkungen zwischen externen Phänomenen und neuronalen Prozessen zu bestimmten Verhaltensregelmäßigkeiten führen, die von Ökonomen mit Hilfe mathematischer Nutzenmodelle reduziert beschrieben werden: Diese Reduktion müßte dann die Funktion des Verhaltens im Kontext des einbettenden Systems abbilden. Diese Beschreibung selbst aber nimmt nicht mehr auf kausale Prozesse in der ‚black box‘ des Individuums Bezug, sondern – hier ganz im Sinne der ‚wohlwollenden‘ Kritiker der Neuroökonomik (vgl. *Ross* 2007, 2008) – in reduzierter Form auf das Verhalten ökonomischer Agenten im Kontext ökonomischer Systeme, d.h. bezieht sich gar nicht mehr auf konkrete Akteure, sondern auf Konstrukte, die emergente Phänomene auf der Systemebene erfassen. So abstrakt dies klingen mag, so einfach ist es zu interpretieren: Ähnliche Argumente sind klassisch bekannt von *Alchian* (1977) und erneut *Friedman* (1953), nämlich dass über die Mechanismen ökonomischer Systeme angenommen wird, dass sie im Ergebnis ein Akteursverhalten generieren, das den Annahmen der Nutzentheorie entspricht, ohne dass gleichzeitig unterstellt werden muß, dass die Nutzentheorie die Akteure selbst beschreibt: Das berühmte-berühmte ‚as if‘ Argument.

Und genau an dieser Stelle bricht aber mit Macht die institutionelle Perspektive hervor: Denn die Mechanismen ökonomischer Systeme sind selbst institutionelle Phänomene, nämlich Regeln, die das Verhalten ökonomischer Akteure bestimmen. Das ‚as if‘ Argument kann in dieser Sichtweise sehr leicht mit modernen Konzepten rekonstruiert werden, vor allem im Kontext der experimentellen Ökonomik, hier eher im Sinne der Simulation ökonomischer Systeme (klassisch sind etwa die Untersuchungen zur Simulation von Wettbewerbsgleichgewichten mit ‚zero intelligence agents‘, denen dann im obigen Sinne Optimierungsverhalten zugeschrieben werden kann, vgl. *Smith* 2003). Dann erscheint der ökonomische Agent als nichts anderes als ein emergentes Phänomen in ‚Complex Adaptive Systems‘ ist, in denen Regeln auf zwei Ebenen zusammenspielen. Erstens, es gibt Regeln nach denen das System operiert, und zweitens, die ökonomischen Agenten repräsentieren

unterschiedliche Regeln individuellen Verhaltens, die in einem evolutorischen Prozeß selektiert werden. Im Ergebnis werden solche Regeln auf der zweiten Ebene selektiert, die sich durch die ökonomische Nutzentheorie reduziert beschreiben lassen. Das bedeutet wiederum, dass im evolvierenden System viele Varianten koexistieren können, wenn dieses selbst ständig Veränderungen unterworfen ist, und dass die Dominanz bestimmter Regeln der zweiten Ebene durch die Regeln auf der ersten Ebene bestimmt wird. Mit anderen Worten, die analytische Reduktion der Nutzentheorie ist selbst kontingent mit Bezug auf die Regeln, nach denen das System operiert.

Dieser Ansatz ist konsequent externalistisch. Er ist keineswegs neu, sondern steht in der Tradition bedeutender Ansätze in der klassischen Ökonomik und der Soziologie. *Karl Marx* beispielsweise sah den Kapitalisten nicht als naturgegebenes Phänomen an, sondern als Reflex der Gesetzmäßigkeiten der Akkumulation im kapitalistischen System. *Max Weber* begriff ‚Rationalität‘ als Ergebnis des Zusammenspiels ideologischer und institutioneller Entwicklungen und damit den Kapitalismus als historisch kontingentes Phänomen. Alle diese Ansätze sind externalistisch mit Blick auf die Erklärung des individuellen Verhaltens.

Daraus ergibt sich aber, dass die theoretische Korrespondenz der Anwendung von Neurowissenschaften in der Ökonomik gerade nicht die konventionelle Nutzentheorie und damit das ökonomische Standardmodell ist, sondern die Institutionenökonomik (*Herrmann-Pillath* 2009). Hierzu gibt es einen Vorläufer, der von der etablierten Neuroökonomik meines Wissens bislang völlig ignoriert wird: *Veblen*. Allerdings können wir noch weiter zurückgehen, nämlich zu *Smith*. *Adam Smiths* Ökonomik ist konsequent naturalistisch, indem er menschliches Verhalten als ein Zusammenspiel von ‚natürlichen‘ Bestimmungsgründen und historisch gegebenen Institutionen erklärt (vgl. *Ashraf et al.* 2005; *Evensky* 2005). Dabei wird der menschliche Akteur gerade nicht durch ein ökonomisches Standardmodell beschrieben, sondern als ein komplexes System von widerstreitenden inneren Kräften, den Emotionen und kulturell geformten Verhaltensneigungen, das die essentielle Eigenschaft der Selbstreflexion aufweist (in der Figur des ‚spectator‘). Gerade diese Selbstreflexion ist mit der institutionellen Entwicklung konkreter menschlicher Gesellschaften verkoppelt, insofern sie jede Art von Verhalten moralisch durchformt. Der eigeninteressierte und optimierende Akteur des ökonomischen Standardmodells erscheint hier als Reflex eines bestimmten Entwicklungsstadiums der menschlichen Zivilisation, nämlich der bürgerlichen Gesellschaft (vgl. ähnlich *Elias* 1969). Ökonomische Rationalität ist selbst eine institutionelle Form. In diesem Sinne ist *Smith* der bedeutendste Vorläufer eines nicht-reduktionistischen naturalistischen Forschungsprogramms in der Wirtschaftswissenschaften, in dem die Neurowissenschaften und die Institutionenökonomik in einem externalistischen Paradigma zu einer Synthese zusammengeführt werden. Diese Wendung in der Diskussion um die Neuroökonomik möchte ich im Folgenden detaillieren.

B. Kausale Interdependenzen zwischen neuronalen Prozessen und Institutionen: Überblick und externalistische Interpretation

Es gibt drei Perspektiven, aus denen die Neurowissenschaften für die Institutionenökonomik von Bedeutung sind. Umgekehrt muß aber immer auch gefragt werden, wie Institutionen im Zusammenspiel mit neuronalen Prozessen Verhalten determinieren, das selbst nie vollständig auf neuronale Prozesse reduzierbar ist. Beide Denkbewegungen zusammen konstituieren die nicht-reduktionistische Synthese von Neurowissenschaften und Institutionenökonomik:

- Erstens: Wie tragen neuronale Prozesse zur Schaffung und Verankerung von Institutionen bei? Gibt es bestimmte verallgemeinerbare Merkmale der institutionellen Evolution, die wesentlich mit neuronalen Prozessen zusammenhängen?
- Zweitens: Welche Wirkung haben Institutionen auf neuronale Prozesse? Wie wird dadurch der Effekt neuronaler Prozesse auf das Verhalten bestimmt?
- Drittens: Welche Konsequenzen ergeben sich aus diesen beiden Perspektiven für die Gestaltung von Institutionen unter normativen Gesichtspunkten?

Ich möchte mich auf die beiden ersten Perspektiven konzentrieren, aber eingangs den dritten Punkt abklären. Die Neuroökonomik nach *Glimcher* ist eine reine Theorie der Wahlhandlung. Sie blendet damit ausdrücklich Wohlfahrtsbetrachtungen aus (*Glimcher* 2009: 509). Für diese Differenzierung zwischen der Analyse der Wahlhandlungen und der Wohlfahrtstheorie gibt es eine umfassende neurowissenschaftliche Begründung, die auch für die Behandlung anderer Fragestellungen von großer Bedeutung ist. Deswegen müssen wir diesen Punkt zuerst erörtern.

I. Neurowissenschaftliche Grundlagen der Dissoziation von Wahlhandlung und Wohlfahrt

Bereits die psychologische und die verhaltensökonomische Forschung hat gezeigt, dass das Nutzenkonzept ausdifferenziert werden muß. So haben *Kahnemann et al.* (1997) früh auf der Basis empirischer Befunde vorgeschlagen, dass zwischen ‚experience utility‘ und ‚decision utility‘ zu unterscheiden ist. Diese Unterscheidung hat sich in weiteren neurowissenschaftlichen Forschungen bestätigt. Sie hängt mit der Rolle des Lernens im menschlichen Entscheidungsverhalten zusammen. Menschliche Wahlhandlungen sind wesentlich dadurch getrieben, die Ergebnisse von Entscheidungen zu verbessern. *Glimcher* (2009, 2011) hat dieses System auf einige sparsame Hypothesen reduziert, in dem Erwartungen eine zentrale Rolle spielen. Das bedeutet, dass Entscheidungen auf die Differenz zwischen Erwartungen und Ergebnissen reagieren, d.h. dass einer Option desto höherer Nutzen zugeschrieben wird, je größer die positive Abweichung zu den Erwartungen ist, relativ zu einem Status quo (dem Referenzpunkt). Neurophysiologisch und neuroanatomisch ist dieser Prozeß durch die Dopamin-gesteuerten neuronalen Strukturen repräsentiert (*Schultz* 2009). Das bedeutet, bei positiven Abweichungen steigt die Ausschüttung des Neurotransmitters Dopamin, und bei zunehmender Konvergenz zwischen Erwartungen und Ergebnis sinkt sie. Dieser Mechanismus liegt vor allem dem Lernen durch Konditionierung zugrunde und findet sich deshalb nicht nur bei Menschen, sondern etwa auch bei Ratten (deren Verhalten sich demzufolge auch leicht ökonomisch erklären läßt, siehe *Phillips et al.* 2007). Dopamin trägt außerdem zum allgemeinen Wohlbefinden bei.

Nun ist aber das Dopamin-System nur indirekt mit den anderen biologischen Steuerungsmechanismen verbunden, die Zielerfüllung und –abweichung messen. Ein einfaches Beispiel ist die Nahrungsaufnahme. Der sich auftuende Dualismus im neuronalen System wurde unter anderem als Differenzierung zwischen ‚Wanting‘ und ‚Liking system‘ strukturell spezifiziert (*Camerer* 2006; *Berridge* 2009). Das heisst beispielsweise, dass beim menschlichen Essverhalten zwischen dem Wunsch nach und der Entscheidung für ein bestimmtes Nahrungsmittel einerseits, und der sich tatsächlich einstellenden Bedürfnisbefriedigung andererseits unterschieden werden muß. Es ist wichtig zu betonen, dass die beiden Systeme nicht mit der Unterscheidung bewußter und unbewußter Prozesse parallel laufen. Beide Systeme funktionieren grundsätzlich ohne Einbeziehung des Bewußtseins, aber bestimmte Prozesse können bewußt werden. Bei der Nahrungsmittelaufnahme ist also zwischen der Wirkung des Essens auf physiologische Steuerkreisläufe, die Zielerreichung signalisieren (etwa: Sättigung), und jenen Mechanismen

zu differenzieren, die auf die Auswahl von Nahrungsmitteln einwirken (*Finlayson et al. 2010*). Bei letzteren handelt es sich um die Dopamin-Mechanismen. Diese Unterscheidung ist wesentlich, um systematische Dysfunktionen zu erklären (*Ross et al. 2007*). Sie können immer dann auftreten, wenn der Dopamin-Prozeß als Ergebnis einer individuellen Lerngeschichte teilweise von den physiologischen Steuerkreisläufen entkoppelt ist. Dann ist es beispielsweise möglich, dass aufgrund vergangener Erfahrungen die Wahrnehmung eines Hamburgers mit positiven Erwartungen assoziiert ist, und damit zu wiederholten Esshandlungen führt, gleichzeitig aber die getroffenen Entscheidungen nicht die physiologischen Steuerungskriterien erfüllen. Im Extremfall ist der übermäßige Konsum von Hamburgern im Moment der Entscheidung subjektiv positiv belegt, führt aber nicht nur zu Übergewicht, sondern schon etwas später zu Unzufriedenheit mit dem resultierenden körperlichen Zustand: ‚Wanting‘ impliziert nicht notwendig ‚Liking‘. Damit wird aber ein zentrales Element des ökonomischen Standardmodells unterhöhlt, das die in Wahlhandlungen offenbarten Präferenzen auch als nicht hintergehbaren Indikator der individuellen Wohlfahrt betrachtet.

Die Differenz zwischen ‚Wanting‘ und ‚Liking‘ ist von zentraler Bedeutung für Wohlfahrtsbetrachtungen, und steht auch im Mittelpunkt der öffentlichen Wahrnehmung der Neuroökonomik insbesondere in den USA. Einer der führenden Verhaltensökonomien, der sich auch intensiv mit der Neuroökonomik befasst, hat sie zu einem Dreh- und Angelpunkt der normativen Konsequenzen erklärt (*Camerer 2006*). Wenn nämlich das ‚Wanting system‘ und das ‚Liking System‘ systematisch auseinanderdriften können, bedeutet das, dass es selbst in einer konsequent individualistischen Betrachtungsweise ein stringentes Argument für staatliche Intervention gibt (*McFadden 2007*). Denn beide Systeme beziehen sich ausschließlich auf das Individuum. Wenn aber das ‚Wanting System‘ zu Handlungsergebnissen führt, die aus der Perspektive des ‚Liking Systems‘ nachteilig sind, bedeutet dies, dass subjektiver Nutzen und individuelle Wohlfahrt auseinanderfallen. Eine externe Korrektur der Wahlhandlungen würde einen Beitrag zur Steigerung der individuellen Wohlfahrt leisten.

Daraus ergeben sich weitreichende Konsequenzen für die Gestaltung von Institutionen, die von *Thaler and Sunstein (2009)* als ‚paternalistic libertarianism‘ bezeichnet wurden. Kerngedanke ist, dass Institutionen oder auch einfach Mengen von Optionen stets so gestaltet werden müssen, dass sie Verhaltensoptionen in einer Weise präsentieren, dass ‚Liking‘ und ‚Wanting‘ konvergieren. Diese Strategie wird mit dem Wort ‚Nudging‘ beschrieben: Beispielsweise könnte eine staatliche Regulierung von Speisekarten in Restaurants gewährleisten, dass die Speisen so arrangiert sind, dass diejenigen mit höherer Wahrscheinlichkeit frei gewählt werden (‚Wanting‘), die auch die vorzugswürdigeren im Sinne des ‚Liking‘ sind. Andere wichtige Beispiele sind Pensionsmodelle, die zwischen ‚opt-in‘ und ‚opt-out‘ Optionen unterscheiden. Für einen rationalen Akteur sollte dies keinen Unterschied machen. Wenn aber Akteure hyperbolische Zeitpräferenzen haben und ‚loss aversion‘ besitzen, ist die Wahrscheinlichkeit wesentlich höher bei einem ‚opt out‘ System, dass sie eine ausreichende Alterssicherung wählen, als bei einem ‚opt in‘ system.

Nun sind diese Beispiele nicht notwendig auf die Neuroökonomik bezogen. Wie bereits ausgeführt, sind die meisten der Ansätze bereits psychologisch und verhaltensökonomisch ausreichend fundiert, und oft bringt die neurowissenschaftliche Sicht eigentlich nur eine zusätzliche empirische Validierung. So oder so sind die Konsequenzen für die institutionelle Gestaltung unter Wohlfahrts Gesichtspunkten weitreichend, weil sie die Souveränität des Individuums relativieren, in Bezug auf die *eigenen* Wohlfahrtskriterien optimale Entscheidungen zu treffen.

Auf der anderen Seite ergibt sich die Schlußfolgerung nicht zwingend, dass der Staat korrigierend eingreifen müsse. Eine alternative Konzeption stellt die Frage, ob und wie weit

die institutionelle Evolution nicht langfristig solche Dilemmata auflöst. Diese Frage ist von der ökonomischen Literatur längst im Zusammenhang der ‚Commitment‘ Mechanismen diskutiert worden (*Gul und Pesendorfer 2001; Bénabou und Tirole 2004; Benhabib und Bisin 2005*). Institutionen können unter bestimmten Bedingungen als ‚Commitment‘ Mechanismen aufgefasst werden, vor allem, wenn ein weiterer Institutionenbegriff verwendet wird, der auch informale Institutionen einschließt. Ein einfaches Beispiel ist die Tatsache, dass Institutionen oft Entscheidungstypen zusammenfassen, gleichermaßen diachronisch und synchronisch (*Ainslie 2007*). Beispielsweise ist ein akademischer Abschluß eine Institution: Sie fasst synchronisch ein bestimmtes Wissensgebiet zusammen, und diachronisch einen Prozeß, der eine gewissen Zahl von Jahren dauert. Dies Institution trägt unter anderem dazu bei, ein Problem der Selbstbindung zu lösen, das bestehen würde, wenn wir das Wissen sozusagen ‚just in time‘ erwerben würden. Bei einem solchen hypothetischen Modell würde sich bei jeder Einzelentscheidung der Effekt hyperbolischer Zeitpräferenzen einstellen: Das Individuum würde weniger lernen, als es eigentlich will, weil es ständig der Verführung kurzfristig scheinbar vorteilhafterer Alternativen verfällt. Wenn aber ein Mal die Grundsatzentscheidung gefällt wurde, ein Studium anzutreten, entsteht ein starker Selbstbindungsmechanismus, der diese Effekte zumindestens teilweise ausgleicht, und der durch starke externe Sanktionen stabilisiert wird. Die Institution des Studiums konstituiert ein ‚Wanting‘, das auch dem ‚Liking‘ im Sinne einer Maximierung der individuellen Wohlfahrt im gesamten Lebenszyklus entspricht.

Ohne Zweifel haben viele Entscheidungen diesen Charakter, insbesondere mit Blick auf den gesamten Lebenszyklus. Dysfunktionen sind umgekehrt auch oft Ausdruck impliziter Institutionalisierungen. Beispielsweise kann Alkoholismus als Folge der Akkumulation von Einzelentscheidungen im Zeitablauf interpretiert werden, in denen Faktoren wie hyperbolische Zeitpräferenzen oder Dopamin-basierte neuronale Prozesse wirksam sind. Die Gewohnheit selbst bildet quasi-institutionelle Merkmale aus, wie etwa das Trinken in bestimmten Umgebungen. Umgekehrt verursacht genau dieser Umstand auch die Schwierigkeiten, von der Sucht wieder los zu kommen, und ein entscheidender Faktor ist eine ‚Um-Institutionalisierung‘. Beispielsweise spielt es eine entscheidende Rolle, ein neues Selbstbild durch institutionalisierte Verhaltensmuster zu etablieren (*West 2006: 137*).

Die Frage stellt sich also, wie weit gerade durch institutionelle Evolution Konstellationen individuellen Verhaltens entstehen, die das Dilemma von ‚Wanting‘ versus ‚Liking‘ auflösen. Genau das ist aber eine konsequent externalistische Position, und sie ist in der Wirtschaftswissenschaft bestens durch das Werk von *Hayeks* repräsentiert. In seiner Theorie der Institutionen vertritt Hayek die Auffassung, dass Institutionen selbst Wissen inkorporieren, das Individuen nicht vollständig zugänglich ist, aber gleichwohl die Grundlage individueller Entscheidungen darstellt (vgl. *Herrmann-Pillath 2010b*). *Hayeks* Theorie ist konsequent individualistisch, externalisiert aber die Wissensbasis individueller Wahlhandlungen in wesentlicher Hinsicht. Auf diese Weise kann die neurowissenschaftliche Dissoziation zwischen Wahl und Wohlfahrt aufgehoben werden. Isoliert betrachtet, wäre das Individuum nicht in der Lage, die Entscheidungen zu treffen, die optimal unter dem Gesichtspunkt der individuellen Wohlfahrt sind; gerade wenn aber Verhalten institutionell gebunden ist, und dabei nicht mehr individuell disponierbar wird, werden die individuellen Optima erreichbar. Wir werden im Folgenden sehen, dass diese Einsicht in der Tat wesentlich für die Integration von Neurowissenschaften und Institutionenökonomik ist. *Hayeks* Theorie folgend (siehe auch *Hayek 1979* zu den ‚drei Quellen menschlicher Werte‘), sind für das menschliche Individuum Prozesse konstitutiv, die auf beiden Ebenen der Population und der körperlichen Einheit stattfinden.

II. ‚Einer Institution entsprechend handeln‘: naturalistische Zugänge

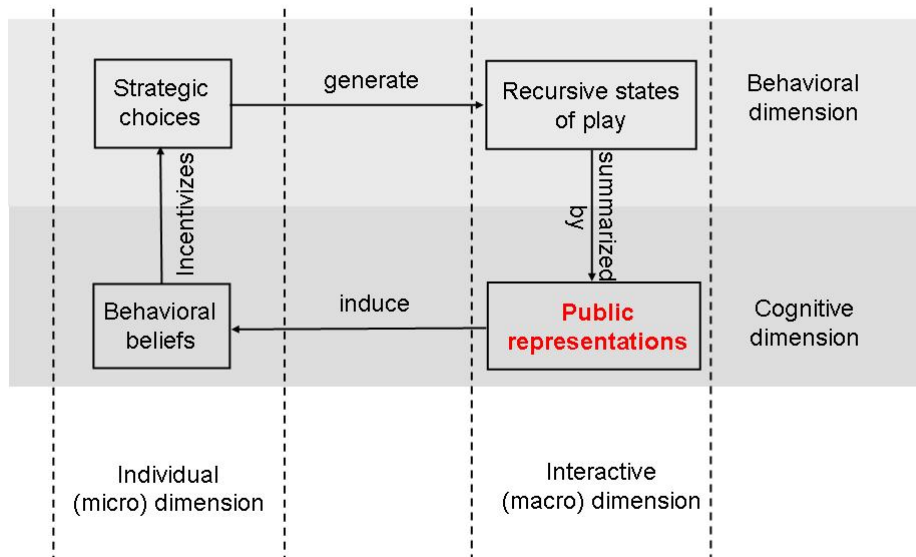
Ich wende mich nun der ersten Frage zu. Wie sind Institutionen neuronal verankert? Hierzu ist es erforderlich, einen eigenständigen theoretischen Rahmen zu skizzieren (*Herrmann-Pillath* 2010a, 2011). Der Standardansatz in der Ökonomie ist weitgehend mentalistisch und reduziert dabei Institutionen auf Anreizstrukturen (z.B. *Smit et al.* 2011). Das letztere bedeutet, dass eine Institution im Grunde als Arrangement von Anreizen betrachtet wird, die bestimmtes Verhalten erzeugen. Die Erzeugung selbst wird als rationale Entscheidung begriffen. Rationalität gilt bei Vertragstheorien auch für die Entstehung von Institutionen, d.h. zumindestens im Prinzip gilt für alle Institutionen, dass sie rational reduzierbar sind, gleichermaßen bezüglich ihrer Entstehung, ihrer Erhaltung und ihrer Wirkung. ‚Mentalismus‘ bezeichnet die Annahme, dass Institutionen letzten Endes gegenseitige Verhaltenserwartungen sind, die durch Handlungen unterschiedlicher Individuen erzeugt werden, d.h. Institutionen koordinieren mentale Zustände. Das wird besonders in spieltheoretischen Konzeptionen deutlich, wo es auch eine gewisse Ambiguität gibt, was die Interpretationen von Institutionen als den ‚Spielregeln‘ versus den Gleichgewichten, also den Ergebnissen des Spiels unter diesen Spielregeln anbetrifft (vgl. etwa *Aoki* 2001: 24ff.; *Dixit* 2004: 5ff.). Mentalismus bedeutet, dass externe Tatbestände keinen weiteren Einfluß auf das Spiel haben. Das Problem der mentalistischen Position ist, dass alle Zustände inferentiell auf der Basis kognitiv zugänglicher Informationen generiert werden. Es gibt spieltheoretisch natürlich die Alternative, evolutionäre Modelle zu verwenden. Diese sind aber in dem Sinne inkompatibel mit der ökonomischen Standard-Theorie, weil sie Individuen direkt als inkorporierte Regeln interpretieren, also als unterschiedliche Strategien, die nicht selbst Gegenstand rationaler Wahl mehr sind. Diese Modellierung entspricht kongenial dem *Hayekschen* Modell der Regelselektion.

Aoki (2007, 2011) hat nun eine Modifikation des spieltheoretischen Grundmodells vorgeschlagen, die direkt auf eine neuroökonomische Deutung hin angelegt ist. Sie baut sich folgendermaßen auf. Unterschieden werden zwei Dimensionen mit zwei Ebenen. In der einen wird die Mikroebene von der Makroebene differenziert. Die Mikroebene betrifft die individuellen Zustände und Entscheidungen, die Makroebene die aggregierten Verhaltensmuster in einer Population von Akteuren. In der anderen wird die Dimension des Verhaltens von derjenigen kognitiver Zustände differenziert. Werden beide Perspektiven in einem Diagramm aufeinander projiziert, ergibt sich folgendes Bild vom institutionellen Prozeß (siehe Abbildung 1). Zunächst wird standardmäßig davon ausgegangen, dass individuelle Strategien den Ausgang von Spielen determinieren, der sich dann auf der Populationsebene als Gleichgewicht abbildet, das sich über die Zeit hinweg reproduziert. Neu kommt nun aber hinzu, dass die Verhaltensebene externe Muster erzeugt, die als ‚Kürzel‘ zur Repräsentation der Gleichgewichtszustände dienen (‚summary representations‘). Diese haben eine Funktion auf der kognitiven Ebene, sind aber selbst ‚öffentlich‘, d.h. extern aus der Sicht der Individuen. Sie erzeugen auf der individuellen Ebene bestimmte Überzeugungen, die dann der Generierung von Strategien zugrundeliegen.

In *Aokis* Modell wird also die kognitive Ebene zerlegt in eine interne und eine externe Seite, wobei die interne individuell und nicht zwischen Individuen geteilt, und die externe kollektiv ist. Ein Beispiel: Als Ergebnis vieler wiederholter Geschäfte eines bestimmten Typs sei ein Vertragsmuster entstanden, das durch eine Handelskammer veröffentlicht und verbreitet wird. Das Vertragsmuster ist ein ‚Kürzel‘ für die vergangenen Spiele. Es ist selber vielen Interpretationen gegenüber offen, kanalisiert aber die Überzeugungen der verschiedenen Akteure, so dass letzten Endes auch der Strategieraum in gewisser Weise eingeschränkt wird. Auf diese Weise wird auch der Kreis möglicher Gleichgewichte eingeschränkt, so dass sich das Muster auf kollektiver Ebene verlässlich reproduziert. Dabei kann das individuelle Wissen über das Vertragsmuster sehr unterschiedlich sein (was *Aokis*

Ansatz grundlegend von vielzitierten Konzept der ‚shared mental models‘ unterscheidet, siehe *Denzau und North 1994*).

Abbildung 1: Aokis institutionentheoretischer Ansatz

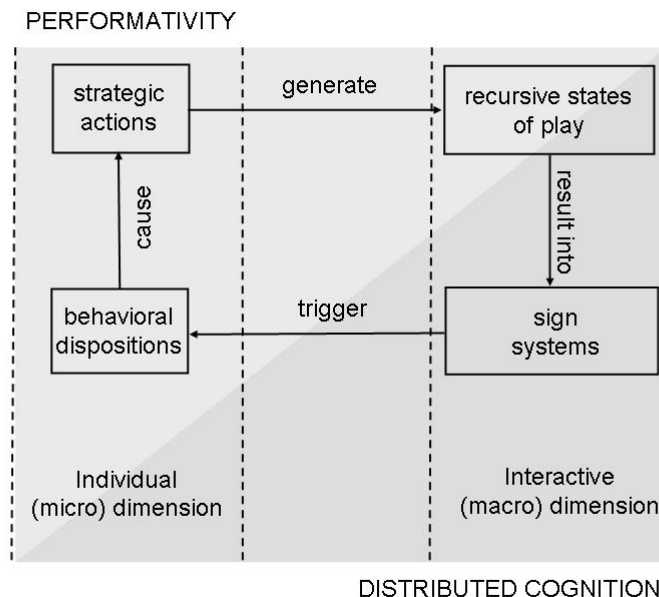


Diese Theorie ist in einem entscheidenden Punkt externalistisch: Sie bezieht strategische Gleichgewichte direkt auf externe Tatbestände, die selbst Ergebnis von Interaktionen der Vergangenheit sind. Welche diese sind, ist fast beliebig, denn letzten Endes geht es um Konventionen. Von zentraler Bedeutung ist aber die Sprache, denn deren Worte und Bedeutungen sind in fundamentaler Weise kollektiv und extern bestimmt, d.h. Phänomene der Populationsebene, gleichzeitig aber individuellen Interpretationen gegenüber offen. Die Sprache spielt nun auch eine entscheidende Rolle in der Institutionentheorie von *Searle* (1995, 2010), die eine direkte Übersetzung der Theorie *Aokis* in die neurowissenschaftliche Perspektive erlaubt. Es ist hier kein Raum, diese Theorie im Detail zu betrachten (für eine komprimierte Zusammenfassung siehe *Searle 2005*). Wichtig ist, dass *Searle* alle Institutionen als sprachliche Phänomene auffasst, in denen ein Bedeutungstransfer stattfindet (seine sogenannte ‚status function‘). Durch diesen Transfer wird ein bestimmter Sachverhalt als ein anderer behandelt, also etwa eine Person ‚als Richter‘ oder ein Stück Papier ‚als Geld‘. Der Transfer geht mit der Schaffung einer Deontologie einher, das heißt bestimmten Rechten und Pflichten. Er ist wesentlich auf die Existenz kollektiver Intentionalität angewiesen, wie sie der Sprache bereits inhärent ist. Ich habe andernorts (*Herrmann-Pillath 2010a*) argumentiert, dass diese Theorie letztendlich eine Theorie der ‚Performativität‘ ist: Performative Sprechakte sind Sprechakte, die eine bestimmte Realität schaffen, wie etwa die Verkündigung eines Urteils. Ähnlich wird eine Institution zum Bestandteil einer sozialen Ontologie, indem ihr sprachlicher Bedeutungstransfer kollektiv anerkannt wird, und die Akteure diesem Verständnis der sozialen Situation folgen.

Für die Brücke zur Neuroökonomik ist nun ausschlaggebend, dass *Searle* das Befolgen von Regeln als eine neuronal verankerte Disposition deutet - das heißt, ähnlich wie eine Fußballspielerin viele regelgebundene Verhaltensweisen internalisiert hat, also eine Disposition besitzt, beim Auftreten bestimmter externer Reize in einer bestimmten Weise zu handeln. Genau hier liegt auch die Brücke zu *Aoki*: Externe Reize sind öffentliche

Repräsentationen, und sie erzeugen bestimmte individuelle Zustände. Die Differenz besteht aber darin, dass diese Zustände keine mentalen sind, also ‚beliefs‘, sondern Dispositionen, d.h. neuronale Zustände. Diese Zustände erzeugen wiederum Strategien als offenbartes Verhalten, die mental repräsentiert sein können, aber nicht müssen. Die Beziehung zwischen Dispositionen und Verhalten ist selbst stochastisch, d.h. im Wesen einer Disposition liegt es, dass es Varianz im offenbarten Verhalten geben kann. Auf der Populationsebene werden dann die Spielgleichgewichte konventionell erzeugt. Dort gilt auch wieder *Aokis* Argument, dass Spielgleichgewichte in öffentlichen Repräsentationen Niederschlag finden, die aus diesen generiert werden. Ich bezeichne diese allgemein als ‚Zeichen‘ bzw. Zeichensysteme‘, die auf die Gleichgewichte Bezug nehmen (vgl. *Skyrms* 2004, 2010). Die Sprache ist ein solches Zeichensystem, aber beispielsweise auch Symbole, die in Architektur verkörpert sind.

Abbildung 2: Institutionen: Performativität und verteilte Kognition



Ich fasse diese an *Searle* anknüpfende, aber auch über ihn hinausgehende Sicht auf Institutionen in Abb. 2 zusammen, die das Schema von *Aoki* in wesentlicher Hinsicht modifiziert. Dort wird die Mikro-Makro-Unterscheidung aufgehoben und durch eine querverlaufende Differenzierung von ‚Performativität‘ und ‚verteilter Kognition‘ ersetzt. Der Begriff der verteilten Kognition bezieht sich auf die jüngeren Ansätze in den Kognitionswissenschaften, menschliche Kognition nicht auf das Gehirn alleine zu beziehen, sondern auf kausale Interaktionen zwischen Artefakten und anderen Individuen mit dem Gehirn, dessen kognitive Leistungen durch diese Verteilung potenziert bzw. oft erst ermöglicht werden (*Hutchins* 1995, 2005; *Sterelny* 2004; siehe die Rezeption bei *North* 2005). Beispielsweise ermöglicht die Schrift andere kognitive Leistungen als die gesprochene Sprache. Diese kommen aber selbstverständlich erst durch die Verknüpfung mit neuronalen Zuständen zum Tragen. Im Ergebnis bedeutet das aber, dass die Mikro-Makro-Differenzierung für die kognitive Ebene hinfällig ist: Die neuronalen Zustände sind zwar eindeutig ontologisch individualisiert, aber die Artefakte sind Phänomene der Populationsebene, da sie keine ausschließlich individuelle Geltung besitzen können. Insofern integrieren die kognitiven Prozesse die Mikro- und die Makroebene. Umgekehrt wird die Mikro-Makro-Differenzierung auch auf der Verhaltensebene obsolet. Denn das individuelle

Verhalten reflektiert aufgrund der Kanalisierung durch die verteilte Kognition bereits kollektive Phänomene. Ein konkretes Beispiel ist die Rolle der Imitation im sozialen Lernen. Diese Mikro-Makro-Integration schlägt sich im Phänomen der Performativität nieder: Die individuelle Strategie gründet in kollektiver Intentionalität im Sinne *Searles* und konvergiert also zur Spiellösung auf der Populationsebene. Das individuelle Befolgen einer Institution ist ein performativer Akt in einer Gemeinschaft von Verwendern der symbolischen Ressourcen, die die Institution konstituieren.

Dieses Modell von Institutionen ist neuroökonomisch, aber geht über die Neurowissenschaft auch hinaus. Der neuroökonomische Aspekt kommt durch die zentrale Bedeutung neuronaler Zustände in Kausalketten zum Tragen, die nicht mental repräsentiert sind (aber sein können, und zwar im Falle der Reflexivität). Jedoch ist die derzeit praktizierte Neuroökonomik individualistisch, d.h. fokussiert ausschließlich Zustände individueller Gehirne. Die Theorie verteilter Kognition stellt dies grundsätzlich in Frage (vgl. programmatisch zu eine ‚social neureconomics‘ *Oullier et al. 2008*, und mit Blick auf die Neuroökonomik *Wilcox 2008*). Dieser Übergang bildet die Spannung zwischen Externalismus und Internalismus ab: Die etablierte Neuroökonomik ist theoretisch inkonsistent internalistisch, während die hier vorgeschlagene Theorie externalistisch ist (vgl. ausführlich hierzu *Herrmann-Pillath 2011b*). Mit anderen Worten, gerade ein neurowissenschaftlich fundierter Ansatz zum Phänomen ‚einer Institution folgen‘ führt zur Konsequenz, die Institution selbst als ein nicht-neuronal konstituiertes Phänomen zu betrachten. Das ist die Grundidee der sozialen Ontologie *Searles*: Sie ist naturalisch im doppelten Sinne, nämlich Individuen als neuronale Strukturen zu begreifen, und Institutionen selbst als eine andere Art von Tatbeständen.

Ein Beispiel für diese Auffassung ist das Geld. Verhaltensökonomisch ebenso wie neuroökonomisch konnte inzwischen nachgewiesen werden, dass die Geldillusion ein psychologisches Phänomen ist, das in alternativen Gehirnzuständen abgebildet wird (*Shafir et al. 1997; Weber et al. 2009*). Dies kann so interpretiert werden, dass die Institution selbst, die in Artefakten repräsentiert wird, neuronale Dispositionen generiert, die eine Neigung zur Geldverwendung systematisch induzieren (z.B. durch das ‚Vertrauen‘ in das Geld). Die Verwendung einer bestimmten Währung als ein institutionelles Gleichgewicht ist also neuronal verankert.

Die vorgeschlagene neuroökonomische Institutionentheorie kann verallgemeinert werden durch die neuronale Theorie der Metapher (*Lakoff 2008*). Das entspricht im Wesentlichen der Position *Searles*, spezifiziert aber die konkreten neuronale Mechanismen wesentlich genauer. Der Grundgedanke besteht darin, dass Regelbefolgung durch bestimmte neuronale Muster unterstützt wird, die sensorimotorisch sind, und dabei die Wertigkeit der Handlungen im weitesten Sinne emotional fundieren, konkret etwa in Gestalt von ‚somatic markers‘ (*Bechara und Damasio 2005*). Diese Theorie besitzt direkte Verbindungen zu Theorien der Empathie und des sozialen Lernens (*Singer und Lamm 2009*). Im Kern geht es darum, dass unsere kognitiven Strukturen zurückgeführt werden auf aggregierte und abstrahierte sensomotorische Schaltkreise, die Perzeption und Aktion direkt verkoppeln (*Fogassi 2011*). Die Aktion kann dabei genauso gut die eigene Aktion sein wie die beobachtbare Aktion anderer, die über sogenannte ‚Spiegelsysteme‘ oder ‚Spiegelneuronen‘ den eigenen Aktionen neuronal homolog sind, d.h. meine eigene Wahrnehmung, dass ein anderer Mensch schwimmt, ist verkoppelt mit meiner eigenen, sozusagen „neuro-virtuellen“ Schwimmhandlung. Die Wurzel dieser Verkopplung liegt in der Ontogenese, also den Mechanismen des kindlichen Lernens. Sie wird gesteuert durch Wertungen, die selbst neuronal verankert sind (*Hurley 2008*). Ein bestimmtes sensorimotorisches Muster ist also über einen ‚somatic marker‘ verankert und wird entsprechend generiert, wenn der externe Reiz zur Aktivierung des ‚somatic marker‘ auftritt.

Für die Analyse von Institutionen bedeutet dies, dass erstens, Regelbefolgung selbst ein sensorimotorisches Muster ist, das bestimmte Handlungen über die Selbstwahrnehmung mit vergangenen Handlungen und deren Wertigkeit verkoppelt, und dass zweitens, eine direkte Kopplung mit beobachteter Regelbefolgung durch andere erfolgt. Die neuronale Theorie der Metapher erlaubt es, diesen Zusammenhang direkt auf die Ebene der Zeichensysteme zu übertragen. Das bedeutet, Institutionen sind, im Sinne *Searles*, auf Metaphern basiert, die selbst wiederum auf sensorimotorischen Muster zurückgehen, die durch ‚somatic markers‘ verankert sind. Um erneut das Beispiel des Geldes zu nehmen: Zum einen wird die Institution getragen von der Sensorimotorik des Geldausgebens und –verwendens, und zwar im ganz konkreten Sinne des Umgangs mit Geld, der die Beobachtung sozialer Praktiken mit der Selbstbeobachtung verkoppelt. Zum anderen ist Geld mit ‚somatic markers‘ verknüpft (die über die neuroökonomischen Nachweis der Geldillusion sogar direkt beobachtbar sind): Ein Vorschlag besteht darin, ähnlichen Argumentationsmustern in der Evolutionären Psychologie folgend (*Tooby und Cosmides 2005*), dass das Artefakt Geld ‚somatic markers‘ aktiviert, die mit dem spezifisch menschlichen Instinkt der sozialen Reziprozität verbunden sind (vgl. *Lea und Webley 2006*). Geld ist also nicht nur ein Tauschmedium, sondern eine direkte Repräsentation der Metapher für sozialen Tausch. Diese neuronal verankerte Metapher erklärt dann auch wieder das Phänomen der Geldillusion, d.h. also dass dem Artefakt selbst Wert beigemessen wird.

Abschließend möchte ich in diesem Abschnitt auf die Bedeutung allgemeiner kognitiver Regelmäßigkeiten hinweisen, die ihrerseits in der spezifischen Funktionsweise des menschlichen Gehirns gründen. Dieser Gedanke ist in der Institutionenökonomik prononciert durch *Schlicht (1998)* und *Kubon-Gilke (1997)* entwickelt worden und geht auf die Gestaltpsychologie zurück. Institutionen können dann als ‚Gestalten‘ aufgefasst werden, die Regelmäßigkeiten im Verhalten von Individuen, insbesondere im Vergleich zu anderen Individuen, repräsentieren. Beispielsweise gäbe es in einer Population eine bestimmte Häufigkeit, mit der Individuen öffentlich auf die Strasse spucken. Die Institutionalisierung des Nicht-Spuckens wird erheblich beschleunigt, wenn die Individuen bei kleineren Verschiebungen in der beobachtbaren Häufigkeit dieses Verhaltens den Referenzrahmen wechseln, ähnlich einem Necker-Würfel; d.h. selbst marginale Veränderungen können durch kognitive Effekte einen Gestalt-Wechsel erzeugen, durch den der ‚normale‘ Zustand anders definiert wird, also vom ‚Spucken‘ zum ‚Nicht-Spucken‘. Ist aber das Nicht-Spucken der normale Zustand, verändern die Gestalteffekte – in Verbindung mit vorher diskutierten neuronalen Mechanismen der Imitation – die individuellen Verhaltensdispositionen, so dass Spucken innerhalb kürzester Zeit aus dem Strassenbild verschwinden kann. Diese Zusammenhänge können komplex sein, machen aber auch deutlich, dass selbst politische Eingriffe anders begründet werden müssen. Ein Beispiel ist das Rauchen. Es gibt inzwischen neurowissenschaftliche Befunde, dass die skizzierten Imitationseffekte direkt das Rauchverhalten beeinflussen (*Wagner et al. 2011*). Dann spielt es tatsächlich eine große Rolle, wie weit das Rauchen als beobachtbare Handlung aus der Öffentlichkeit ausgeschlossen wird, also über die Bedeutung des ‚second hand smoking‘ hinaus.

Ich fasse die Ergebnisse dieses Abschnittes dahingehend zusammen, dass wir heute *Hayeks* Sicht auf Institutionen wesentlich genauer neurowissenschaftlich fundieren können. Anders als in vielen ökonomischen, insbesondere spieltheoretischen Konzeptionen, sind Institutionen nicht lediglich ein Reflex der Koordination mentaler Zustände individuell optimierender Akteure. Institutionen wurzeln in neuronal inkorporierten Handlungsdispositionen von Individuen, die mit externen Symbolsystemen kausal korrelieren. Ein solcher Ansatz läßt sich auch in die spieltheoretischen Modelle reintegrieren, etwa in Form der Bestimmungsgründe von ‚salience‘ bei der Erklärung der Emergenz von Konventionen (*Sugden 2011*). Entscheidend ist die Einsicht, dass das Befolgen von Institutionen nicht selbst ein rationaler

Akt im Sinne des ökonomischen Standardmodells sein kann. Institutionen sind gerade in einem neurowissenschaftlich fundierten Ansatz irreduzibel. Wenn dem so ist, stellt sich die Frage, wie die kausalen Effekte von Institutionen analysiert werden können.

III. Generische Wirkungen von Institutionen: Ein konzeptioneller Rahmen

Bei der Diskussion der Wirkungen von Institutionen erscheint ein Aspekt als der zentrale, den vor allem *Don Ross* (2005, 2007) herausgearbeitet hat. *Ross* geht seinerseits von einem Vorschlag von *Binmore* (1994) aus, bei der Analyse sozialer Strukturen und Interaktionen zwei Ebenen zu unterscheiden. Die eine Ebene ist diejenige strategischer Interaktionen, die mit dem konventionellen spieltheoretischen Instrumentarium analysiert werden kann. Die andere Ebene ist das sogenannte ‚game of life‘, in dem bestimmte Merkmale von Akteuren als Ergebnis eines evolutischen Prozesses determiniert werden, d.h. letztendlich als biologische Merkmale der Art ‚Mensch‘. *Ross* schlägt nun vor, eine dritte Ebene zwischen diesen beiden einzuschalten, auf der die Identitäten der Akteure bestimmt werden. Mit anderen Worten, es gibt allgemeine biologische Merkmale von Menschen, und darüber hinaus bestimmte persönliche Merkmale, die ihre Identität in sozialen Interaktionen ausmachen. Das entspricht Ansätzen in der Spieltheorie, ‚Typen‘ von Spielern durch bestimmte Signalsysteme zu identifizieren, und verallgemeinert betrachtet, geht es letzten Endes darum, wie sich ein ‚common knowledge‘ über Spieler konstituieren kann.

Dabei sieht *Ross* die menschliche Sprache als Schlüssel an: Sie ist ein Problem, das sich gleichzeitig seine Lösung schafft. Sie ist ein Problem, weil mit der Entstehung der Sprache die Möglichkeit entstand, beliebige kontrafaktische Sachverhalte zu kommunizieren, insbesondere auch, über die eigenen Absichten und Motive falsche Informationen zu verbreiten. Sprache hat daher die Achillesferse, in ein ‚cheap talk‘ zu degenerieren, in dem letzten Endes gar keine Information mehr übertragen wird. Auf der anderen Seite ist Sprache aber in grundlegender Weise ein intersubjektives Phänomen und erlegt daher den Sprachverwendern bestimmte Restriktionen auf. Beispielsweise lassen sich die Bedeutungen von Worten nicht beliebig manipulieren, oder gibt es logische bzw. allgemeinere Kriterien der Konsistenz von sprachlichen Äußerungen. *Ross* argumentiert daher, dass Sprache bestimmte Zwänge erzeugt, die eigenen ‚narratives‘ selbst als Datum zu behandeln. Die ‚narratives‘ über die eigene Person, die im kollektiven Gedächtnis sozialer Interaktionen akkumulieren, werden dann konstitutiv für die persönliche Identität der Akteure. Jeder Spieler wird als geschichtliche Entität konstituiert, und dieser Geschichte kann er sich nicht mehr in beliebiger Weise entziehen. Damit erhält aber auch die Sprache ihre kommunikative Funktion, kann also nicht mehr zum ‚cheap talk‘ implodieren.

Ich habe an anderer Stelle ausführlich dargestellt (*Herrmann-Pillath* (2010/2011)), dass dieser Ansatz enge Bezüge zur Diskussion des Identitätskonzeptes in der Wirtschaftswissenschaft aufweist (*Akerlof* und *Kranton* 2000; *Davis* 2003; *Kirman* und *Teschl* 2006). Wie *Davis* (2007) gezeigt hat, ist es erforderlich, zwischen sozialer und personaler Identität zu unterscheiden. Die soziale Identität entsteht durch die Zuweisung sozialer Klassifizierungen (wie etwa Geschlechterrollen), die personale Identität bildet den Anker, der diese verschiedenen Klassifizierungen zusammenhält. Wie dies geschieht, macht *Ross* plausibel: Die personale Identität ist die sprachlich vermittelte geschichtliche Natur des Individuums. Da auch alle sozialen Klassifikationen sprachliche Kategorien sind, können wir auch sagen, dass die beide Dimensionen der sozialen und personalen Identität nach Synchronie und Diachronie differenziert sind: Die soziale Identität entsteht durch die soziale Klassifizierung zu einem bestimmten Zeitpunkt, die personale Identität bestimmt, wie die derzeitige Position zustande gekommen ist, und in welche Richtung sie sich weiter entwickeln wird (siehe *Horst* et al. 2006).

Diese Thematik ist nun in unserem Zusammenhang von besonderem Interesse, weil die Frage der Identität sich keineswegs nur für *Alter* in sozialen Interaktionen stellt, sondern auch für *Ego* selbst! Wir selbst wollen wissen, wer wir sind, und oft ist unser Wissen darüber unvollständig. Warum ist das so? Die moderne Gehirnwissenschaft gibt hier eine Antwort, die inzwischen auch in die Wirtschaftswissenschaft Eingang gefunden hat (*Brocas* und *Carrillo* 2008): Das Gehirn ist kein vollständig integriertes Entscheidungssystem, sondern modularisiert, dynamisch und komplex. Das bedeutet zunächst, dass es hochgradig instabil sein könnte. Wo kommt die Konstanz des Akteurs her, wie sie sich in dessen Identität insbesondere diachronisch niederschlägt? Im Lichte der modernen Gehirnwissenschaft ist die Annahme der Wirtschaftswissenschaft begründungsbedürftig, dass Präferenzen stabil sind.

Diese Fragestellung ist selbst strittig. Im kanonischen Ansatz *Glimchers* wird die Eigenschaft der ökonomischen Nutzenfunktion in die Neurowissenschaft importiert, ein konsistentes und integriertes System abzubilden. Viele Ergebnisse der Gehirnwissenschaft und Psychologie legen aber ein ‚multiple selves‘ Modell nahe, das selbst evolutionär erklärt werden kann. Die Evolutionäre Psychologie beispielsweise geht davon aus, dass das Gehirn unterschiedliche Module aufweist, die auf unterschiedliche Entscheidungsprobleme hin angepasst sind (*Tooby* und *Cosmides* 2005). Das Gehirn nutzt also sozusagen die Vorteile der Arbeitsteilung aus. Die verschiedenen Module werden durch Emotionen integriert, und nicht durch übergreifende Entscheidungsmechanismen. Auf noch grundlegender Ebene ist das Gehirn selbst als ein dynamisches evolvierendes System zu begreifen, das ständig Wahrnehmungen und Verhalten aus einem endogenen Selektionsprozeß heraus generiert, in dem es keine zentrale Entscheidungsinstanz gibt (*Dennett* 1995; *Edelman* 1987, 2006; *Oullier* und *Kelso* 2009). Daraus entsteht aber das Problem, wie dieses hochgradig komplexe und dynamische System den Absturz in Chaos vermeidet. *Edelman* (1987) hat die Lösung bereits sehr früh vorgeschlagen, dass dies nur durch externe Verankerung möglich ist, nämlich durch Kommunikation mit anderen Gehirnen. Auf den Menschen gemünzt bedeutet das: Nur durch die Sprache kann das menschliche Gehirn stabile Strukturen ausbilden, durch die sich dann Akteure formieren. Das Gehirn gewinnt seine Identität durch externe Anker.

Diese These muß mit ‚multiple selves‘ Ansätzen zusammengesehen werden. In solchen Ansätzen entsteht das Problem, dass menschliche Handlungen grundsätzlich konfliktbehaftet sind. Beispielsweise lassen sich intertemporale Entscheidungsprobleme als Konflikte zwischen heuten und zukünftigen ‚selves‘ deuten (*Jamison* und *Wegener* 2010). Hier ist kein Raum, auf die Fülle von konzeptionellen Alternativen einzugehen, sondern ich konzentriere mich auf einen zentralen Punkt: In jedem ‚multiple selves‘ Modell gibt es ein internes Informationsproblem des modularisierten Gehirns. Das heisst, die unterschiedlichen Module sind unvollständig über andere Module informiert. Wenn wir dann davon ausgehen, dass ein Modul Verhalten generiert, muß dieses Verhalten also nicht optimal sein mit Blick auf andere Module. Die einfachste Form eines solchen Modells ist bereits früh vom Soziologen *James Coleman* (1990) vorgeschlagen worden, nämlich zwischen ‚object self‘ und ‚acting self‘ zu unterscheiden, die jeweils einer evaluativen und einer Handlungsinstanz entsprechen. Damit gelangen wir also zu den bereits diskutierten Ansätzen zurück, ‚Liking‘ und ‚Wanting‘ zu differenzieren. Dann können wir also sagen, dass das Handlungssystem nicht vollständig informiert ist über das Bewertungssystem. Es entsteht ein Principal-Agent problem (*Brocas* und *Carrillo* 2008).

Wie kann das Handlungssystem wissen, was das Bewertungssystem präferiert? Ich schlage vor, dass die Identität einer Person diese Information bietet. Das ist ein ungemein radikaler Schritt, der aber erstaunlicherweise direkt in den Kern des ökonomischen Standardmodells zurückführt. Wie ich eingangs betont habe, ist das ökonomische Standardmodell eigentlich externalistisch, d.h. baut auf der Theorie offener Präferenzen auf. Wird es à la *Glimcher* internalistisch fehlinterpretiert, würde das Gehirn also tatsächlich als Optimierer begriffen.

Diese Interpretation ist mit einiger Sicherheit empirisch falsch. Wenn man am Externalismus des Standardansatzes festhalten will, muß vielmehr postuliert werden, dass auch die Präferenzen selbst externe Phänomene sind, also gerade nicht ‚subjektiv‘: Die offenbaren Präferenzen sind in gar keiner Weise ‚hintergebar‘. Wie ist das möglich? Das ist möglich, wenn das Handlungssystem die Informationen über das Bewertungssystem dadurch gewinnt, dass es Informationen über realisierte Handlungen und deren Ergebnisse auswertet. Die Geschichte von Wahlhandlungen, reflektiert in sprachlich vermittelten Narrativen, *ist* die Präferenzstruktur (vgl. ähnlich *Ariely* und *Norton* 2007). Erst die Identität des Akteurs schafft die Integration des komplexen Systems Gehirn, und diese Identität konstituiert sich im Sinne von *Ross* als öffentliche ‚narrative‘, die *Ego* nicht nur anderen, sondern auch sich selbst erzählt..

Damit gelangen wir aber zu unserer eigentlichen Frage zurück, wie Institutionen wirken. Es gibt hier höchst interessante Bezüge zur ‚alten‘ Institutionenökonomik (*Herrmann-Pillath* 2009). *Veblen* hatte die Auffassung entwickelt, dass Institutionen ‚habits of thought‘ sind, also in wesentlicher Weise internalisiert. Auch in modernen Interpretationen werden Bezüge zwischen ‚habits‘ und Institutionen hergestellt (*Hodgson* 2004). Das gerade vorgetragene Argument knüpft genau hier an: Präferenzen könnten als ‚habits‘ gedeutet werden, nämlich als erlernte Muster von Wahlhandlungen, die ihre Stabilität durch ihre Geschichtlichkeit erhalten. Das ist weniger weit entfernt vom ökonomischen Standardansatz als es anmuten mag, denn eigentlich bereits in *Beckers* (1996) Ansatz angelegt, der zwischen biologischen Fundamentalpräferenzen und erlernten ‚tastes‘ unterscheidet. *Sobel* (2005) weist auch darauf hin, dass *Beckers* Modell formal homolog zu ökonomischen Modellen von Identität ist, also eine wesentliche soziale Komponente einführt.

Dann können wir also die Hypothese formulieren, dass die generische Wirkung von Institutionen darin besteht, die personale Identität des Akteurs herzustellen und zu stabilisieren. Der Mensch bzw. sein Gehirn sind in fundamentaler Weise auf Institutionen angewiesen. Ohne Institutionen verliert er seinen Charakter als Akteur, ja, er kann noch nicht einmal in den ‚Naturzustand‘ zurückfallen, weil sein Gehirn die direkten Reiz-Reaktionsmechanismen dynamisch entkoppelt hat.

In diesem Sinne können wir also so weit gehen, den Präferenzen des Standardmodells selbst den Charakter von Institutionen zuzuschreiben. Genau das liegt aber den Mechanismen zugrunde, die ich bereits in Abschnitt I diskutiert habe. Betrachten wir erneut das Beispiel des Studiums. Ein ‚multiple selves‘ System ist damit konfrontiert, dass es ständig intertemporale Entscheidungskonflikte generiert, und außerdem regelmäßig in suboptimale Zustände einmünden würde. Die Lösung besteht darin, bestimmte Sequenzen von Entscheidungen zusammenzufassen. Genau das leistet aber die Institutionalisierung, etwa eines akademischen Abschlusses. Die Entscheidung besteht nicht mehr darin, jeden Tag zu überlegen, ob weiter Wissen erworben werden soll, sondern nurmehr darin, sich für ein bestimmtes Studium zu entscheiden. Das ist eine Entscheidung über künftige Identitäten. Ist sie getroffen, sind wesentliche Handlungsmuster institutionell vorgeprägt. Dadurch wird aber das intertemporale Entscheidungsproblem entschärft: Was zu präferieren ist, wird externalisiert.

Ich denke, dass alle spezifischeren Wirkungen von Institutionen mit dieser fundamentalen Rolle als Identitätsstifter zusammenhängen. Das betrifft vor allem die indirekten Mechanismen, die über die Imitation wirken. Insofern Institutionen ähnliche Verhaltensweisen in einer Population von Akteuren unterstützen, verstärken sie die internen Bindungen durch die Effekte von Imitation. In diesem Sinne sind hier also personale und soziale Identität direkt miteinander verkoppelt. Denn die Imitation setzt nicht direkt an sozialen Klassifikationen an, sondern lediglich an beobachtbaren Verhaltensmustern. Diese verstärken dann aber Verhaltensähnlichkeiten in einer Population, die unmittelbar in personalen Identitäten verankert sind, gleichzeitig aber soziale Klassifikationen bestätigen

und akzentuieren. In diesen Zusammenhang sind Institutionen auch kognitiv relevant, weil sie selbst Gestaltbildende Kraft haben. Jede Institution ist auch ein Interpretationsrahmen, unterstützt also sogenannte ‚frames‘. ‚Frames‘ werden inzwischen als wesentliches Element realistischer ökonomischer Modelle der Wahlhandlung angesehen (*Gintis* 2006).

Hier ist kein Raum, auf ‚frames‘ im Detail einzugehen, das Konzept findet verschiedene Ausprägungen und ist keineswegs eindeutig definiert, insbesondere auch weil es andere Begriffe gibt, deren gegenseitige Abgrenzung oft unklar ist. In unserem Zusammenhang möchte ich die These formulieren, dass es sich um ein fundamentales und generisches Phänomen handelt, das in der kognitiven Autonomie der Gehirnprozesse begründet ist: Menschliche Kognition enthält immer ein wesentliches Element der Interpretation, d.h. es gibt keine ‚objektive‘ Referenz zur Außenwelt, sondern nur relativ zu den evolvierten kognitiven Strukturen, die ihrerseits neuronal inkorporiert sind (diese Position ist grundlegend entwickelt in *Hayek* 1952, siehe auch *Herrmann-Pillath* 1992). Bei dieser Position muß aber wieder ein internalistischer Fehlschluß vermieden werden: Denn solche Interpretationen sind ohne externen Anker beliebig. Dem klassischen *Wittgensteinschen* (1958) Argument gegen die Möglichkeit einer ‚Privatsprache‘ folgend (Überblick bei *Candlish* 2004), müssen wir vielmehr davon ausgehen, dass alle Bedeutungen nur über soziale Interaktionen und im weitesten Sinne sprachliche Kommunikation konstituiert werden. ‚Frames‘ sind also nur in dem Sinne neuronal inkorporierte kognitive Strukturen, als die neuronale Inkorporation wesentlich für die betrachteten Kausalprozesse ist. Diese Prozesse selbst sind aber externalistisch zu analysieren.

Diese Perspektive ist für die Analyse von Institutionen wichtig, weil sie ermöglicht, Institutionen selbst als ‚frames‘ zu betrachten. In der neueren ökonomischen Literatur taucht diese Dimension gewöhnlich in der Behandlung der ‚Kultur‘ als Determinante von Verhalten auf. Dabei wird aber ein vergleichsweise undifferenzierter Begriff der ‚Kultur‘ verwendet, der faktisch lediglich auf die eigentlich willkürliche Abgrenzung von Populationen Bezug nimmt (also die ‚Kulturen‘ von Nationalstaaten oder von ethischen Gruppierungen) (zum Überblick etwa *Guiso et al.* 2006, zur Kritik *Herrmann-Pillath* 2010c). Bei genauerer Betrachtung geht es bei der Kultur jedoch um die konkrete Rolle von Institutionen als Interpretationsrahmen. Die inzwischen klassische Analyse hierzu ist die Untersuchung von *Henrich et al.* (2005) zu den kulturellen Unterschieden im Verhalten bei einperiodigen Ultimatumspielen: Bei der vergleichenden Analyse von 15 verschiedenen ethnischen Gruppierungen erwies sich, dass das Anbieterverhalten ebenso wie die Reaktionen der Empfänger von Teilungsvorschlägen sehr unterschiedlich ist und in jedem Fall klar von der Vorhersagen des ökonomischen Standardmodells abweicht. Die Erklärung besteht vermutlich darin, dass das Spiel selbst als eine Institution interpretiert wird. Das bedeutet, die Spielregeln werden im Lichte ähnlicher Interaktionen gedeutet, die in diesen Populationen in der Vergangenheit stattgefunden haben. Wenn also beispielsweise Ultimatumspiele in einer Population von Walfängern durchgeführt werden, wo die Teilung des Fangs eine etablierte Institution ist, werden sehr hohe Angebote in Ultimatumspielen beobachtet. Diese Wechselwirkung zwischen vorhandenen Institutionen und Exekution des Spieles ist ein bekanntes Problem in der Verhaltensökonomik: Beispielsweise werden Vertrauensspiele (‚trust games‘) von Probanden oft als Wetten und damit als Glücksspiele interpretiert.

Ein wichtiges Ergebnis der Untersuchung von *Henrich et al.* ist, dass – vielleicht kontraintuitiv? – die Integration von Populationen in weiter reichende Marktbeziehungen die Bereitschaft zum Teilen erhöht. Damit bestätigt die Untersuchung die *Smithsche* Position, dass Märkte selbst institutionell fundiert sind, und sie ermöglichende institutionelle Evolution wesentliche moralische Dimensionen aufweist. Das ‚ökonomisch rationale‘ Verhalten tritt vor allem dann auf, wenn die Populationen Lebensformen aufweisen, in denen weiter reichende Interaktionen keine Bedeutung besitzen, und die ökonomische Subsistenz durch die engere

Familie gesichert wird. Anders gesagt, der ‚homo oeconomicus‘ erscheint selbst als ein institutionelles Phänomen.

Diese Thematik führt zum allgemeinen Problem der ‚sozialen Präferenzen‘, d.h. dem empirisch gesicherten Tatbestand, dass Menschen in hohem Maße die Präferenzen anderer berücksichtigen, etwa durch Fairness-Normen, aber vor allem auch durch altruistische Sanktionen, d.h. die Bereitstellung von Sanktionen unbeteiligter Dritter (*Hopfensitz* und *Reuben* 2009). Letzteres Phänomen ist auch wichtig für die Stabilisierung von Institutionen: Soziale Präferenzen generieren offenbar positive Gratifikationen, die aus der Möglichkeit von Sozialität herrühren, eine Idee, die *Adam Smith* bereits mit dem Begriff des ‚fellow feeling‘ begründet hat (*Sugden* 2002). Daher werden ähnliche neuronale Mechanismen aktiviert, die auch andere Entscheidungen über die Wahl von Handlungsalternativen bestimmen, d.h. die Möglichkeit von Kooperation oder die Möglichkeit der Durchsetzung von Kooperation aktiviert die dopaminergischen Strukturen im Gehirn (zum Überblick *Fehr* 2009). Letzten Endes führt dies auch zum grundlegenden Konzept der *Smithschen* Moralphilosophie zurück, der ‚sympathy‘. In den Neurowissenschaften wird *Smith* daher zunehmend auch als Vorläufer moderner Theorien der Empathie betrachtet (*Singer* und *Lamm* 2009).

Nun hat aber *Smith* ebenfalls schon gezeigt, dass die Möglichkeit der Empathie selbst auf Institutionen angewiesen ist. *Smiths* ‚sympathy‘ ist zwar eine universelle menschliche Fähigkeit, aber sie wird nur konkret in der Interaktion zwischen Individuen in spezifischen sozialen Kontexten. Gelebte ‚sympathy‘ ist erlernte ‚sympathy‘ und als solche kontingent. Daher besteht für die *Smithsche* Theorie auch das Problem, letzten Endes moralische Normen selbst nur als konkrete Institutionen in konkreten historischen Konstellationen begründen zu können (*Forman-Barzilai* 2010). Dieses Problem vermittelt aber in unserem Zusammenhang tatsächlich eine wesentliche Einsicht: Soziale Präferenzen sind auch in dem Sinne ‚sozial‘, als sie ohne eine institutionelle Einbettung inhaltsleer bleiben. Interessanterweise hatte die Untersuchung von *Henrich* et al. auch gezeigt, dass individuelle Variation (etwa Ausbildung, Geschlecht oder Wohlstand) für die Variation des Verhaltens im Ultimatumspiel nicht erheblich ist, sondern nur die Variation über ‚kulturelle‘ Merkmale der Populationen.

Institutionen ermöglichen soziale Präferenzen vor allem durch die kognitive Reduktionsleistung. *Ego* kann *Alter* ‚verstehen‘, weil es geteilte Institutionen und Erwartungen bezüglich deren Sanktionierung gibt. Die Institution ermöglicht Wissen über *Alter*, und gleichzeitig gibt sie *Ego* Orientierung über sich selbst. Das bekannte Konzept der Komplexitätsreduktion durch Institutionen kann damit unmittelbar neurowissenschaftlich begründet werden bzw. es wird möglich, diese Reduktion direkt mit bestimmten Erkenntnissen über die Funktionsweise des Gehirns zu begründen. Im Vergleich zum reduktionistischen Ansatz von *Glimcher* kann angemerkt werden, dass *Glimcher* selbst herausarbeitet, wie zentral Referenzpunkteffekte für elementare Wahlhandlungen sind. Solche Referenzpunkte sind ‚frames‘, und insofern liegt genau hier der entscheidende Schwachpunkt seines kanonischen Modells. Im Lichte der hier vorgetragenen Überlegungen sind Referenzpunkte wesentlich extern konditioniert und angewiesen auf externe Anker. Das bedeutet aber auch umgekehrt, dass nur im Rahmen eines konsequent externalistischen Ansatzes die Neuroökonomik überhaupt erst konsistent formuliert werden kann.

An dieser Stelle wäre noch vieles zu sagen, doch möchte ich mit diesen Betrachtungen über die generischen Wirkungen von Institutionen begnügen. Diese wurzeln in der spezifischen Dynamik von Performativität und verteilter Kognition, die ich im vorherigen Abschnitt bereits dargestellt habe. Hinzu kommt nun das Konzept der Modularität des dynamischen Systems Gehirn und die Idee, dass Institutionen konstitutiv für seine Identität als Akteur sind.

C. Schluß: Individuen und Institutionen als Elemente einer naturalistischen sozialen Ontologie

In diesem Beitrag habe ich versucht zu zeigen, dass Institutionenökonomik und Neuroökonomik natürliche Partner im Geschäft sind, menschliches Verhalten zu erklären. Diese These stellt sich in kritische Distanz zum derzeit führenden kanonischen Modell der Neuroökonomik, das von *Glimcher* entwickelt wurde, und greift damit auch die kritische Tradition der Institutionenökonomik klassischer Prägung auf.

Es ist in diesem Zusammenhang erstaunlich zu notieren, wie weitgehend *Adam Smith* eine solche Synthese bereits antizipiert hatte. Im Gegensatz zur langen Geschichte der Fehlinterpretationen von *Smith* als Theoretiker des rationalen Eigennutzes hatte *Smith* eine Konzeption des Menschen vorgelegt, die naturalistisch war, im Sinne einer Analyse der ‚human nature‘. Diese menschliche Natur sah *Smith* als hochkomplex an: Das Individuum ist ein dynamisches System von Trieben, Emotionen und reflexiven Mechanismen, das essentiell auf soziale Interaktion angewiesen ist, um zu sinnvollem und zweckadäquatem Handeln gelangen zu können. Wahlverhalten ist Ergebnis von Lernprozessen in sozialen Kontexten, und daher gerade in der reflektierten Form institutionell geprägt. Reflektion wird erst durch Institutionen ermöglicht, die es erlauben, die im Prinzip unüberwindbare physikalische Barriere zum Denken der Anderen zu überwinden. *Smiths* ‚spectator‘, moralisch vervollkommnet als ‚impartial spectator‘, ist Ausdruck dieser wesentlichen Kopplung interner Prozesse und externer Institutionen.

Diese Sichtweise steht deshalb in der kritischen Tradition der Institutionenökonomik, weil sie im Sinne *Searles* eine andere soziale Ontologie begründet als diejenige der Standardökonomik. *Glimchers* Internalismus ergibt sich zwingend aus der unreflektierten Übernahme dieser sozialen Ontologie: Die soziale Realität besteht ausschließlich aus Individuen, letztendlich definiert durch deren Körperlichkeit. Die klassische Institutionenökonomik hatte hingegen Institutionen stets als Gegenstände eigener Art behandelt. *Searle* steht in dieser Tradition, indem er zwischen unterschiedlichen Arten von ‚facts‘ unterscheidet: In *Searle* (1995) als ‚observer independent‘ versus ‚observer relative facts‘, in *Searle* (2010) ersetzt er den Term ‚observer‘ durch ‚intentionality‘ oder ‚mind‘. Institutionen sind keine Gegenstände wie sie gewöhnlich in physikalischen Theorien betrachtet werden, wie etwa ein Stein oder ein Neuron. Beide existieren unabhängig vom menschlichen Beobachter. Eine Institution ist hingegen ein Fakt, der nur relativ zum menschlichen Beobachter existiert. Der Geldschein ist ein physikalisches Objekt, wird aber nur zum Geld durch die Interpretation. Diese ist aber selbst nicht beliebig, denn die Interpretation gründet im Phänomen der ‚kollektiven Intentionalität‘, wie sie in Reinform in der menschlichen Sprache auftritt. Das einzelne Individuum kann zwar autonome Interpretationen entwickeln, aber um ‚Geld‘ als Fakt zu etablieren, ist kollektives Verhalten erforderlich, das in einer gemeinsamen interpretativen Haltung wurzelt. Ist Geld aber als Fakt konstituiert, erhält es eigenständige kausale Kraft, die nicht mehr im Belieben der Individuen steht, und die gerade deshalb auch Gegenstand beispielsweise makroökonomischer und ökonomischer Analysen werden kann.

Eine analytisch angemessene soziale Ontologie enthält also Individuen und Institutionen. Ich habe versucht zu zeigen, dass diese Sicht heutzutage empirisch durch die Entwicklung der Neurowissenschaften gestützt werden kann. Das menschliche Gehirn ist ein hochkomplexes System, das ohne externe Anker in überbordernder Dynamik kollabieren würde. Institutionen sind dieser Anker. Das Schlüsselkonzept ist dasjenige der ‚Identität‘. Nur Institutionen ermöglichen die Stabilität und Konstanz von Akteuren in der Zeit. Obgleich sie Phänomene der Populationsebene sind, werden sie konstitutiv für *individuelle* Identität. Das Individuum selbst ist also ein institutionelles Phänomen. *Davis* (2003, 2006) hat das so formuliert, dass

personale Identität nur durch eine deontologische Komponente möglich wird. Institutionen schaffen Verbindlichkeiten, die nicht nur für andere wichtig sind, sondern vor allem auch für das Individuum selbst in seiner Fähigkeit zur Reflexion. Ohne Institutionen ist eine Konzeption des ‚Selbst‘ nicht möglich.

Ich möchte abschließend betonen, dass diese soziale Ontologie keineswegs deckungsgleich ist mit der im späten 19. Jahrhundert entstandenen Dissoziation zwischen ‚Individuum‘ und ‚Gesellschaft‘, also auch der entsprechenden Ausdifferenzierung zwischen Ökonomik und Soziologie. Der Begriff der Institution ist zunächst völlig neutral gegenüber der Abgrenzung der Population, in der Institutionen evolvieren. Die minimale Einheit ist die Triade, d.h. zwei Individuen, die sich gegenüber einem beobachtenden Dritten abgrenzen, indem sie einer Institution folgen. Darüber hinaus sind Institutionen in beliebiger Weise kombinierbar mit unterschiedlichen Gruppengrößen und –strukturen.

Die enge konzeptionelle Verkopplung zwischen Neuroökonomik und Institutionenökonomik bedeutet für die Neuroökonomik, dass sie ohne eine Theorie institutioneller Evolution notwendig unvollständig bleiben muß. Es ist prinzipiell unmöglich, beobachtbares Verhalten ausschließlich durch interne neuronale Prozesse zu erklären. Neuronale Prozesse werden erst im Zusammenspiel mit externen institutionellen Fakten kausal spezifiziert. Die Theorie institutioneller Evolution ist aber selbst keine neurowissenschaftliche Theorie, sondern befasst sich mit der interaktiven Dynamik von Individuen und sozialen Aggregaten. *Glimchers* Ansatz interpretiert daher den Begriff der NeuroÖKONOMIK in irreführender Weise: *Ross* (2005, 2008) folgend, führt gerade das ökonomische Element der Neuroökonomik eine Dimension ein, die nicht mehr in Kategorien der Neurowissenschaften greifbar ist. Nur gilt umgekehrt auch, dass eine Theorie des Individuums nur als naturalistische möglich ist, soll sie empirisch gehaltvoll sein. Hierfür sind neurowissenschaftliche Erkenntnisse grundlegend.

Literatur

- Ainslie*, G. (2007): Thought Experiments That Explore Where Controlled Experiments Can't: The Example of Will, in: *Ross et al.* (2007): 169-198.
- Akerlof*, G. A. and *Kranton*, R. E. (2000): Economics and Identity, *Quarterly Journal of Economics* CXV(3): 715-753.
- Aoki*, M. (2001): *Toward a Comparative Institutional Analysis*, Stanford: Stanford University Press.
- Aoki*, M. (2007): Endogenizing Institutions and Institutional Change, *Journal of Institutional Economics* 3(1): 1-31.
- Aoki*, M. (2011): Institutions as cognitive media between strategic interactions and individual beliefs, *Journal of Economic Behavior and Organization* (2011), forthcoming, doi:10.1016/j.jebo.2011.01.025
- Ariely*, D. / *Norton*, M. I. (2007): How Actions Create – Not Just Reveal – Preferences, *Trends in Cognitive Science* 12(1): 13-16.
- Ashraf*, N., *C. F. Camerer* und *G. Loewenstein* (2005): Adam Smith, Behavioral Economist, *Journal of Economic Perspectives* 19(3): 131-146.
- Bechara*, A. / *Damasio*, A. R. (2005): The Somatic Marker Hypothesis: A Neural Theory of Economic Decision, *Games and Economic Behavior* 52: 336-372.
- Becker*, G. S. (1996): *Accounting for Tastes*, Cambridge: Harvard University Press.
- Becker*, W. J. / *Cropanzano*, R. / *Sanfey*, A. G. (2011): Organizational Neuroscience: Taking Organizational Theory Inside the Neural Black Box, *Journal of Management* 37(4): 933-961.

- Bénabou, R./ Tirole, J. (2004): Willpower and Personal Rules, *Journal of Political Economy* 112(4): 848-886.
- Benhabib, J./ Bisin, A. (2005): Modeling Internal Commitment Mechanisms and Self-Control: A Neuroeconomics Approach to Consumption-Saving Decisions, *Games and Economic Behavior* 52: 460-492.
- Bernheim, B. D. (2009): The Psychology and Neurobiology of Judgment and Decision Making: What's in it for Economists?, in: *Glimcher et al. (2009)*: 115-126.
- Berridge, K. C. (2009): ‚Liking‘ and ‚Wanting‘ Food Rewards: Brain Substrates and Roles in Eating Disorders, *Physiology & Behavior* 97: 537-550.
- Binmore, K. (1994): *Game Theory and the Social Contract Volume I: Playing Fair*, Cambridge / London: MIT Press.
- Brocas, I. / Carrillo, J. D. (2008): The Brain as a Hierarchical Organization, *American Economic Review* 98(4): 1312-1347.
- Buss, D. M., ed. (2005): *The Handbook of Evolutionary Psychology*, Hoboken: Wiley.
- Camerer, C. (2006): Wanting, Liking, and Learning: Neuroscience and Paternalism, *The University of Chicago Law Review* 73: 87-110.
- Camerer, C. F. (2007): Neuroeconomics: Using Neuroscience to Make Economic Predictions, *Economic Journal* 117(519): C26-C42.
- Camerer, C. F. / Loewenstein, G. / Prelec, D. (2005): Neuroeconomics: How Neuroscience Can Inform Economics, *Journal of Economic Literature* Vol. XLIII: 9-64.
- Candlish, S. (2004): Private Language, *The Stanford Encyclopedia of Philosophy* (Spring 2004 Edition), Edward N. Zalta (ed.), URL = <<http://plato.stanford.edu/archives/spr2004/entries/private-language/>>.
- Coleman, J. (1990): *Foundations of Social Theory*. Cambridge/London: Belknap.
- Davis, J. B. (2007): Akerlof and Kranton on Identity in Economics: Inverting the Analysis, *Cambridge Journal of Economics* 31(3):349-362.
- Davis, J.B. (2003): *The Theory of the Individual in Economics. Identity and Value*. London and New York: Routledge.
- Dennett, D. C. (1995): *Darwin's Dangerous Idea. Evolution and the Meanings of Life*, New York: Simon & Schuster.
- Denzau, A. T./North, D. C. (1994): Shared Mental Models: Ideologies and Institutions, *Kyklos* 47: 3-32.
- Dixit, A. K. (2004): *Lawlessness and Economics. Alternative Modes of Governance*, Princeton and Oxford: Princeton University Press.
- Edelman, G. M. (1987): *Neural Darwinism. The Theory of Neuronal Group Selection*. New York: Basic Books.
- Edelman, G. M. (2006): *Second Nature. Brain Science and Human Knowledge*. New Haven and London: Yale University Press.
- Elias, N. (1969a/1990): *Über den Prozeß der Zivilisation. Soziogenetische und psychogenetische Untersuchungen*. Erster Band: Wandlungen des Verhaltens in den weltlichen Oberschichten des Abendlandes. Frankfurt: Suhrkamp.
- Evensky, J. (2005): Adam Smith's Theory of Moral Sentiments: On Morals and Why They Matter to a Liberal Society of Free People and Free Markets, *Journal of Economic Perspectives* 19(3): 109-130, 2005.
- Fehr, E. (2009): Social Preferences and the Brain, in: *Glimcher et al. (2009)*: 215-232.
- Finlayson, G./ Halford, J. / King, N. / Blundell, J. (2010): Chapter 7.3. The Regulation of Food Intake in Humans, <http://www.endotext.org/obesity/obesity7.3/obesity7-3.html>

- Fogassi, L. (2011): The mirror neuron system: How cognitive functions emerge from motor organization, *Journal of Economic Behavior & Organization* 77(1): 66-75
- Forman-Barzilai, F. (2010): *Adam Smith and the Circles of Sympathy. Cosmopolitanism and Moral Theory*, Cambridge: Cambridge University Press.
- Friedman, M. (1953): *Essays in Positive Economics*, Chicago: Chicago University Press.
- Gibbs, R. W., ed. (2008): *The Cambridge Handbook of Metaphor and Thought*, New York et al.: Cambridge University Press.
- Gintis, H. (2006): A Framework for the Unification of the Behavioral Sciences, *Behavioral and Brain Sciences* 30: 1-16.
- Glimcher, P. W. (2003): *Decisions, Uncertainty, and the Brain. The Science of Neuroeconomics*, Cambridge / London: MIT Press.
- Glimcher, P. W. (2009): Choice: Towards a Standard Back-pocket Model, in: *Glimcher et al. (2009)*: 503-522.
- Glimcher, P. W. (2011): *Foundations of Neuroeconomic Analysis*, Oxford and New York: Oxford University Press.
- Glimcher, P. W. / Camerer, C. F. / Fehr, E. / Poldrack, R. A., eds. (2009): *Neuroeconomics: Decision Making and the Brain*, Amsterdam et al.: Elsevier.
- Guiso, L. / Sapienza, P. / Zingales, L. (2006): Does Culture Affect Economic Outcomes? *Journal of Economic Perspectives* 20(2): 23-48.
- Gul, F. / Pesendorfer, W. (2001): Temptation and Self-Control, *Econometrica* 69: 1403-1436.
- Gul, F. / Pesendorfer, W. (2008): The Case for Mindless Economics, in: *Caplin, A. / Schotter, A., eds. The Foundations of Positive and Normative Economics*, Oxford: Oxford University Press, 3-41.
- Harrison, G. W. (2008): Neuroeconomics: A Critical Reconsideration, *Economics & Philosophy* 24(3): 303-344.
- Hayek, F. A. von (1952): *The Sensory Order. An Inquiry into the Foundations of Theoretical Psychology*. Chicago: University of Chicago Press.
- Hayek, F. A. von (1979): *Law, Legislation and Liberty. A New Statement of the Liberal Principles of Justice and Political Economy. Volume 3: The Political Order of a Free People*. Chicago: University of Chicago Press.
- Henrich, J. / Boyd, R. / Bowles, S. / Camerer, C. / Fehr, E. / Gintis, H. / McElreath, R. / Alvard, M. / Barr, J. / Ensminger, A. / Hill, K. / Gil-White, F. / Gurven, M. / Marlowe, F. / Patton, J. Q. / Smith, N. and Tracer, D. (2005): 'Economic Man' in Cross-cultural Perspective: Behavioral Experiments in 15 Small-scale Societies, *Behavioral and Brain Sciences*, 28: 795-855.
- Herrmann-Pillath, C. (1992): The Brain, Its Sensory Order and the Evolutionary Concept of Mind. On Hayek's Contribution to Evolutionary Epistemology, in: *Journal for Social and Biological Structures* 15/2, S. 145-187.
- Herrmann-Pillath, C. (2009): Elements of a Neo-Veblenian Theory of the Individual, *Journal of Economic Issues* XLIII(1): 189-214
- Herrmann-Pillath, C. (2010/11): *The Economics of Identity and Creativity. A Cultural Science Approach*, St. Lucia: University of Queensland Press and New Brunswick and London: Transactions.
- Herrmann-Pillath, C. (2010a): A Neurolinguistic Approach to Performativity in Economics, *Journal of Economic Methodology* 17(3): 241-260.
- Herrmann-Pillath, C. (2010b): Dimensionen des Wissens: Ein kognitiv-evolutionärer Ansatz auf der Grundlage von F.A. von Hayeks Theorie der „Sensory Order“, in: Theresia Theurl, Hrsg., *Institutionelle Hintergründe und Konsequenzen von Wissen*, Schriftenreihe des Vereins für Socialpolitik, Berlin: Duncker&Humblot, 51-84.
- Herrmann-Pillath, C. (2010c): What Have We Learnt From Twenty Years of Economic Research Into Culture? in: *International Journal of Cultural Studies* 13(4): 317-335

- Herrmann-Pillath, C. (2011a): Institutions, Distributed Cognition and Agency: Rule-following as Performative Action, forthcoming: *Journal of Economic Methodology*, penultimate version: Frankfurt School Working Paper Series No. 157, SSRN: <http://ssrn.com/abstract=1745437>.
- Herrmann-Pillath, C. (2011b): Towards an Externalist Neuroeconomics: Dual Selves, Signs, and Choice, *R&R Journal of Neuroscience, Psychology and Economics*.
- Hodgson, G. M. (2004): Reclaiming Habit for Institutional Economics, *Journal of Economic Psychology* 25: 651-660.
- Hopfensitz, A. / Reuben, E.(2009): The Importance of Emotions for the Effectiveness of Social Punishment, *Economic Journal* 119(540): 1534-1560.
- Horst, U. / Kirman, A. / Teschl, M. (2006): Changing Identity: The Emergence of Social Groups, GREQAM Document de Travail 2006-51.
- Hubert, Mirja (2010): Does Neuroeconomics Give New Impetus to Economic and Consumer Research? *Journal of Economic Psychology* 31: 812-817.
- Hubert, Mirja / Kenning, Peter (2008): A Current Overview of Consumer Neuroscience, *Journal of Consumer Behavior* 7: 272-292.
- Hurley, S. (2008): The Shared Circuits Model: How Control, Mirroring and Simulation Can Enable Imitation, Deliberation, and Mindreading, *Behavioral and the Brain Sciences* 31(1): 1-21.
- Hutchins, E. (1995): *Cognition in the Wild*, Cambridge and London: MIT Press.
- Hutchins, E. (2005): Material Anchors for Conceptual Blends, *Journal of Pragmatics* 37: 1555-1577.
- Jamison, J. / Wegener, J. (2010): Multiple Selves in Intertemporal Choice, *Journal of Economic Psychology* 31: 832-839.
- Kahnemann, D. / Wakker, P. P. / Sarin, R. (1997): Back to Bentham? Explorations of Experienced Utility, *Quarterly Journal of Economics* 112: 375-405.
- Kirman, A. / Teschl, M. (2006): Searching for Identity in the Capability Space, *Journal of Economic Methodology* 13(3): 299-325.
- Krohn, W. / Küppers, G., ed. (1992): *Emergenz: Die Entstehung von Ordnung, Organisation und Bedeutung*, Frankfurt: Suhrkamp.
- Kubon-Gilke, G. (1997): *Verhaltensbindung und die Evolution ökonomischer Institutionen*, Marburg: Metropolis.
- Lakoff, G. (2008): The Neural Theory of Metaphor, in: *Gibbs* (2008): 17-38.
- Lau, J., Deutsch, M., "Externalism About Mental Content", *The Stanford Encyclopedia of Philosophy* (Fall 2008 Edition), Edward N. Zalta (ed.), URL = <http://plato.stanford.edu/archives/fall2008/entries/content-externalism/>.
- Lea, S. G. / Webley, P. (2006): Money As Tool, Money As Drug: The Biological Psychology of a Strong Incentive, *Behavioral and the Brain Sciences* 29: 161-209.
- Macdonald, G. / Papineau, D., eds. (2006): *Teleosemantics. New Philosophical Essays*, Oxford / New York: Oxford University Press.
- McFadden, D. (2006): Free Markets and Fettered Consumers, *American Economic Review* 96(1): 5-29.
- Millikan, R. (1989): Biosemantics, *Journal of Philosophy* 86: 281-297.
- Neander, K. (2009): Teleological Theories of Mental Content, *The Stanford Encyclopedia of Philosophy* (Winter 2009 Edition), Edward N. Zalta (ed.), URL = <http://plato.stanford.edu/archives/win2009/entries/content-teleological/>.
- North, D. C. (2005): *Understanding the Process of Economic Change*, Princeton and Oxford: Princeton University Press.
- Oullier, O. / Kelso, J.A. S. (2009): Social Coordination, from the Perspective of Coordination Dynamics, in: *Encyclopedia of Complexity and Systems Sciences*, ed. R.A. Meyers, Berlin: Springer.

- Oullier, O. / Kirman, A. P. / Kelso, J.A.S. (2008): The Coordination Dynamics of Economic Decision Making: A Multilevel Approach to Social Neuroeconomics, *IEEE Transactions on Neural Systems and Rehabilitation Engineering* 16(6): 557-571.
- Phillips, P. E. / Walton, M. E. / Jhou, T. C. (2007): Calculating Utility: Preclinical Evidence for Cost-Benefit Analysis by Mesolimbic Dopamine, *Psychopharmacology* 191: 483-495.
- Ross, D. (2005): *Economic Theory and Cognitive Science: Microexplanations*, Cambridge and London: MIT Press.
- Ross, D. (2008): Two Styles of Neuroeconomics, *Economics & Philosophy* 24: 473-483.
- Ross, D. / Sharp, C., / Vuchinich, R. E. / Spurrett, D. (2008): *Midbrain Mutiny. The Picoeconomics and Neuroeconomics of Disordered Gambling*, Cambridge, Mass.: MIT Press.
- Ross, Don (2007): H. sapiens as Ecologically Special: What Does Language Contribute? *Language Studies* 29: 710-731.
- Roth, G. (1992): Kognition: Die Entstehung von Bedeutung im Gehirn, in: *Krohn and Küppers* (1992): 104-133.
- Schantz, R., ed. (2004): *The Externalist Challenge*, Berlin: de Gruyter.
- Schlicht, E. (1998): *On Costum in the Economy*, Oxford: Clarendon.
- Schultz, W. (2009): Midbrain Dopamine Neurons: A Retina of the Reward System? in: *Glimcher et al.* (2009): 323-329.
- Searle, John R. (1995): *The Construction of Social Reality*, New York: Free Press.
- Searle, John R. (2005): What Is an Institution?, *Journal of Institutional Economics* 1(1): 1-22.
- Searle, John R. (2010): *Making the Social World. The Structure of Human Civilization*, Oxford: Oxford University Press.
- Shafir, Edward / Diamond, Peter / Tversky, Amos (1997): Money Illusion, *Quarterly Journal of Economics* CXII(2): 341-374.
- Singer, T. / Lamm, C. (2009): The Social Neuroscience of Empathy, *Annals of the New York Academy of Sciences* 1156: 81-96.
- Skyrms, B. (2004): *The Stag Hunt and the Evolution of Social Structure*, Cambridge et al.: Cambridge University Press.
- Skyrms, B. (2010): *Signals. Evolution, Learning, & Information*, Oxford: Oxford University Press.
- Smit, J.P. / Buekens, F. / du Plessis, S. (2011): What is Money? An Alternative to Searle's Institutional Facts, *Economics & Philosophy* 27(1): 1-22.
- Smith, Vernon L. (2003): Constructivist and Ecological Rationality in Economics, *The American Economic Review* 93(3): 465-508.
- Sobel, J. (2005): Interdependent Preferences and Reciprocity, *Journal of Economic Literature* XLIII(2): 392-436.
- Stanton, A. A. / Day, M. / Welppe, I. M., eds. (2010): *Neuroeconomics and the Firm*, Cheltenham and Northampton: Edward Elgar.
- Sterelny, K. (2004): Externalism, Epistemic Artefacts and The Extended Mind, in: *Schantz* (2004c): 239-254.
- Sugden, R. (2002): Beyond Sympathy and Empathy: Adam Smith's Concept of Fellow-Feeling, *Economics and Philosophy* 18: 63-87.
- Sugden, R. (2011): Saliency, Inductive Reasoning and the Emergence of Conventions, *Journal of Economic Behavior and Organization* 79: 35-47.
- Thaler, R. H. / Sunstein, C. R. (2009): *Nudge. Improving Decisions about Health, Wealth, and Happiness*, London: Penguin.
- Tooby, J./ Cosmides, L. (2005): Conceptual Foundations of Evolutionary Psychology, in: *Buss* (2005): 5-67.

- Wagner, D. et al. (2011): Spontaneous Action Representation in Smokers when Watching Movie Characters Smoke, *Journal of Neuroscience* 31(3): 894-898.
- Weber, B. / Rangel, A. / Wibral, M. / Falk, A. (2009): The Medial Prefrontal Cortex Exhibits Money Illusion, *Proceedings of the National Academy of Sciences* 106(13): 5025-5028.
- West, R. (2006): *Theory of Addiction*, Oxford: Blackwell and Addiction Press.
- Wilcox, Nathaniel T. (2008): Against Simplicity and Cognitive Individualism, *Economics & Philosophy* 24(3): 523-533.
- Wilson, R. A. (2004): *Boundaries of the Mind. The Individual in the Fragile Sciences: Cognition*, Cambridge et al.: Cambridge University Press.
- Wittgenstein, L.(1958): *Philosophical Investigations*, Oxford: Basil Blackwell.