

Beckmann, Klaus

Working Paper

Jon Elster und das Zeitinkonsistenz-Problem

Diskussionspapier, No. 2007-21

Provided in Cooperation with:

Martin Luther University of Halle-Wittenberg, Chair of Economic Ethics

Suggested Citation: Beckmann, Klaus (2007) : Jon Elster und das Zeitinkonsistenz-Problem, Diskussionspapier, No. 2007-21, ISBN 978-3-86010-961-8, Martin-Luther-Universität Halle-Wittenberg, Lehrstuhl für Wirtschaftsethik, Halle (Saale), <https://nbn-resolving.de/urn:nbn:de:gbv:3:2-3390>

This Version is available at:

<https://hdl.handle.net/10419/170296>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.



Klaus Beckmann

Jon Elster und das Zeitinkonsistenz-Problem

Diskussionspapier Nr. 2007-21

des Lehrstuhls für Wirtschaftsethik
an der Martin-Luther-Universität Halle-Wittenberg,
hrsg. von Ingo Pies, Halle 2007

Haftungsausschluss

Diese Diskussionspapiere schaffen eine Plattform, um Diskurse und Lernen zu fördern. Der Herausgeber teilt daher nicht notwendigerweise die in diesen Diskussionspapieren geäußerten Ideen und Ansichten. Die Autoren selbst sind und bleiben verantwortlich für ihre Aussagen.

ISBN 978-3-86010-960-1 (gedruckte Form)
ISBN 978-3-86010-961-8 (elektronische Form)
ISSN 1861-3594 (Printausgabe)
ISSN 1861-3608 (Internetausgabe)

Autorenanschrift

Prof. Dr. Klaus Beckmann

Universität der Bundeswehr Hamburg
Fakultät für Wirtschafts- und Sozialwissenschaften
Professur für Volkswirtschaftslehre, insb. Finanzwissenschaft
Holstenhofweg 85
22043 Hamburg
Tel.: +49(0)40-6541-2844
Email: klaus.beckmann@hsu-hh.de

Korrespondenzanschrift

Prof. Dr. Ingo Pies

Martin-Luther-Universität Halle-Wittenberg
Juristische und Wirtschaftswissenschaftliche Fakultät
Wirtschaftswissenschaftlicher Bereich
Lehrstuhl für Wirtschaftsethik
Große Steinstraße 73
06108 Halle
Tel.: +49 (0) 345 55-23420
Fax: +49(0) 345 55 27385
Email: ingo.pies@wiwi.uni-halle.de

Kurzfassung

Zeitinkonsistenz wird im Allgemeinen grob als das Phänomen verstanden, dass ein Entscheider, der einen bestimmten optimalen Plan gefasst hat, diesen bei einer späteren Reoptimierung revidiert, ohne dass sich unterdessen relevante Unsicherheit aufgelöst hätte. Die wirtschaftspolitische und institutionenökonomische Relevanz des Themas kann man wohl nur schwer überschätzen.

Was Jon Elster am Zeitinkonsistenz-Phänomen fesselt, das ist die Frage nach dem Wesen der Rationalität. Wiederholt wirft er die Frage auf, ob denn eine Person rational genannt werden dürfe, die nicht imstande sei, einen zeitinkonsistenten Plan durchzuhalten und den einmal gefassten Plan zu dem ursprünglichen Frommen nicht gegen sich selbst durchsetzen könne.

In diesem Papier soll ein simples formales Grundmuster herausgearbeitet werden, das die verschiedenen Fälle von Zeitinkonsistenz eint und das sich in eine allgemeine Systematik von Regelbegründungen einordnen lässt. Damit entfernen wir uns ein Stück von Elster, der die formale Analogie durchaus sieht, aber ihre Anwendungen auf intra- und interpersonelle Konflikte für inkommensurabel hält.

Unter Nutzung dieses einfachen Modells wird sodann Jon Elsters Beitrag zur Theorie der Zeitinkonsistenz gewürdigt. In einem ersten Schritt wird Elsters Forderung nach vollkommener Willensstärke als eines Merkmals wohlverstandener Rationalität kritisiert. Dann aber wird das Argument einer „Meta-Meta-Ebene“, auf der Individuen logisch noch vor der Entscheidung über Spielregeln Eigenschaften ihrer Persönlichkeit wählen, als besonderer Beitrag von Jon Elsters zur Theorie der Zeitinkonsistenz gehoben – führt diese Überlegung doch zu dem Schluss, dass die (Fähigkeit zur) Überwindung von Zeitinkonsistenz nicht in jedem Fall wünschenswert sein muss.

Jon Elster und das Zeitinkonsistenz-Problem

Klaus Beckmann

“Up to this point this survey has proceeded as if time were irrelevant for rational choice. In reality, it is almost the other way around. Choice situations in which time does not play a crucial role are rare.”

Jon Elster (1986, S. 10)

“And, of course, regretting either course later on (at least for the moment) is to be expected, for otherwise precommitment would have had no purpose.”

Robert H. Strotz (1955, S. 173)

Der Bundeskanzler¹ begibt sich am Vorabend einer wichtigen Besprechung zu Bett. Damit er unbesorgt ein- und gut durchschlafen kann, stellt er einen besonders schrillen Wecker auf einen frühen Zeitpunkt – mit einem gehörigen Sicherheitspuffer – ein und platziert ihn unerreichbar weit von seinem Bett. Am nächsten Morgen wacht er also zeitig auf. Eine halbe Stunde hätte er eigentlich noch schlummern können – doch hätte er von dieser Möglichkeit tatsächlich Gebrauch gemacht, wäre eine unruhige, schlaflose Nacht die Folge gewesen.

Nun ist der Kanzler wie geplant fit und eilt beschwingten Schrittes zu seinem Termin beim Finanzminister. Der unterbreitet ihm eine neue Idee: Man könnte doch angesichts knapper Kassen einfach einen Lastenausgleich beschließen und 10% der privaten Vermögen konfiszieren. Diese Vermögen seien ja bereits geschaffen, und so könnten deren Besitzer – noch dazu diese bösen Besserverdienenden (BB) – der Besteuerung gar nicht mehr entgehen.

Kanzler und Richtlinienkompetenz sind freilich dagegen. Der Haken an diesem Plan sei, meinen sie, dass die BB dann für die Zukunft weitere solche „Lastenausgleiche“ erwarteten und daher weniger Vermögen bilden und weniger investieren würden. Zum Zeitpunkt der Wiedervereinigung, da hätte es vielleicht geklappt – damals hätte man ja auf die Einzigartigkeit der Lage verweisen und im Gegenzug den Art. 14 GG etwas stärken können –, aber jetzt ... Doch die Sache mit dem Sparen, führt der Kanzler weiter aus, sei überhaupt nicht so dringlich. Nach den laufenden Umfragen habe man ja ohnedies kaum Chancen auf eine Wiederwahl. Außerdem wisse man, welche Schwerpunkte die Gegenseite bei den staatlichen Ausgaben habe. Dann sei es besser, man mache jetzt Schulden und finanziere die eigenen Lieblingsprojekte; die anderen müssten dann halt nach der Regierungsübernahme auf die Finanzierung von Projekten verzichten, die man selber gar nicht wolle.

¹ Dieses Beispiel bezieht sich offensichtlich auf einen Zeitpunkt in der jüngeren deutschen Vergangenheit.

Unser Kanzler verabschiedet sich von dem breit grinsenden Finanzminister und begibt sich in sein Büro. Da das Rauchen in Bundesgebäuden zum Zeitpunkt unserer Geschichte noch nicht völlig verboten ist, will er sich auf diesen vortrefflichen Start in den Tag eine Cohiba gönnen. Doch o Schreck! Plötzlich fällt ihm ein, dass er, um seinen Appetit auf Zigarren zu zügeln, seit Wochen nur noch drei Stück im Büro lagert – und diese sind bereits aufgeraucht. Gerade in so einer besonderen Situation wäre doch bei Licht betrachtet eine Zigarre angemessen – aber *tant pis*.

Bei allen vier Elementen in dieser erfundenen Geschichte geht es um die *Vorteile von Selbstbindungen* (Stellen des Weckers, Garantien gegen konfiskatorische Besteuerung, Verpflichtung zum Budgetausgleich, Verhinderung des eigenen Konsums von Suchtgütern) und die damit verbundene *Versuchung zum Regelbruch*. Es handelt sich hierbei klar um zwei Seiten derselben Medaille, denn eine Selbstbindung macht offenkundig nur dann Sinn, wenn sie eine Handlung ausschließt, die in der Zukunft gewählt würde, aus heutiger Sicht aber nicht optimal erscheint.²

Die im Beispiel genannten Situationen mit vorteilhafter Selbstbindung und dem späteren Wunsch, aus dieser entlassen zu werden, unterscheidet von den geläufigeren Gefangenen-Dilemmata, dass bereits eine *einseitige* Bindung genügt. Der Mechanismus von Selbstbindungen besteht darin, über die Festlegung eigenen Handelns im Voraus das Handeln anderer *indirekt* in eine gewünschte Richtung zu beeinflussen. (Wobei der „Anderer“ auch eine spätere Inkarnation des Entscheiders selbst sein kann – wie im Fall des Weckers und der Zigarren.) Bei einer Regelbindung zur Überwindung eines Gefangenen-Dilemmas dagegen strebt man eine *direkte* Bindung des Anderen an, und die eigene Bindung ist letztlich nichts anderes als der Preis, den man dafür zu zahlen hat.

Im Vorjahr haben sich die *Konzepte der Gesellschaftstheorie* mit Thomas C. Schelling befasst, in dessen Werk die Installation von Selbstbindungen eine Schlüsselrolle spielt (siehe Beckmann 2007a). Die Probleme, welche dabei auftreten, fasst Schelling (1960, S. 28) glänzend zusammen:

1. (These examples) clearly depend on not only incurring a commitment but on communicating persuasively to the other party.
2. It is by no means easy to establish the commitment, nor is it entirely clear to either of the parties concerned just how strong the commitment is.
3. Similar activity may be available to the parties on both sides.
4. The possibility of commitment (...) is by no means equally available.

In Jon Elsters Werk dagegen – insbesondere in seinen früheren Arbeiten – geht es um die andere Seite der Medaille: um die Versuchung zum Bruch der Selbstbindung oder um die *Zeitinkonsistenz*. Diese wird im Allgemeinen grob als das Phänomen verstanden, dass ein Entscheider, der einen bestimmten optimalen Plan gefasst hat, diesen bei einer späteren Reoptimierung revidiert, ohne dass sich unterdessen relevante Unsicherheit aufgelöst hätte. Die wirtschaftspolitische und institutionenökonomische Relevanz des Themas kann man wohl nur schwer überschätzen, hat doch jede private Vertragsbeziehung mit asymmetrischen Lieferpflichten diese Struktur (de Jasay 1989, S. 22 ff.; Beckmann 1998, S. 422 f.; Binmore 2006, S. 164 f.) und besteht darin ein wesentlicher Grund für die Notwendigkeit von Vertrauen (Lahno 1995, S. 74 ff.). Aus Sicht einer wirtschaftspolitischen Planung wird es – wiederum im engen

² Siehe oben das Strotz-Zitat (1955).

Zusammenhang mit der „Vertrauensfrage“ – um die Ausnutzung von Ankündigungseffekten der Politik gehen (Jensen 2007) und aus Sicht des Militärökonomien um nicht weniger als um die Möglichkeit zum Ergreifen der Initiative (Beckmann 2007a).

Jon Elsters Erkenntnisinteresse liegt anderswo. Was ihn am Zeitinkonsistenz-Phänomen fesselt, das ist die Frage nach dem *Wesen der Rationalität*. In seinem frühen Werk wirft Elster (1987, S. 67) die Frage auf, ob denn eine Person rational genannt werden dürfe, die nicht imstande sei, einen zeitinkonsistenten Plan durchzuhalten und den einmal gefassten Plan zu dem ursprünglichen Frommen nicht gegen sich selbst durchsetzen könne.³ Elster hat Zweifel, kommt aber nach meiner Lesart zu einer negativen Antwort.

Hierin vermag ich Elster nicht ganz zu folgen. Rationalität – auch „richtige“, nicht bloß „begrenzte“ – setzt nicht Übermenschlichkeit voraus, verlangt nicht, bei *jeder* Möglichkeit zur Revision eines Plans am gefassten Entschluss festzuhalten, auch wenn dies ein rationaler Entscheider bei *vielen* solchen Möglichkeiten täte. Wenn man vom rationalen Entscheider bei einem Regelbindungsproblem erwarten kann, dieses durch die schiere Kraft des Willens zu überwinden, wieso nicht auch bei den Übrigen? Damit wäre man indes unmittelbar bei der Forderung, ein rationaler Akteur müsse im einstufigen Zwei-Personen-Gefangenen-Dilemma kooperieren. Einem Diktum Ken Binmores⁴ folgend halte ich selbst Defektion in diesem Spiel nachgerade für ein Charakteristikum eines (instrumental) rationalen Entscheiders und die Grundüberzeugung davon (bei korrekter Definition des Spiels) für das Signum des Ökonomen.

Doch es gibt noch einen tieferen Grund: Die Überwindung von Regelbindungsproblemen zu einem Merkmal der Rationalität im Sinne einer normativen Richtschnur für Entscheider zu erheben könnte zum Ideal eines langweiligen stoischen Lebens führen, das ein rationales Individuum auf einer vorgelagerten Entscheidungsebene nicht notwendig einem Leben in regelbrecherischer Sünde vorziehen würde. Und auch für diese Sichtweise kann man Elster (1999; 2007) zu den *Gefühlen* anführen. Insoweit erscheint mir *Elsters Position widersprüchlich*.

In diesem Papier möchte ich zwei Beiträge leisten:

1. Es soll ein simples formales Grundmuster herausgearbeitet werden, das die verschiedenen Fälle von Zeitinkonsistenz eint und das sich in eine allgemeine Systematik von Regelbegründungen einordnen lässt. Damit entfernen wir uns ein Stück von Elster, der die formale Analogie durchaus sieht, aber ihre Anwendungen auf intra- und interpersonelle Konflikte für inkommensurabel hält (Elster 2000, S. 24).
2. Unter Nutzung des einfachen Modells will ich Jon Elsters Beitrag zur Theorie der Zeitinkonsistenz würdigen. In einem ersten Schritt werde ich die oben bereits umrissene Kritik an Elsters Forderung nach vollkommener Willensstärke als eines Merkmals wohlverstandener Rationalität vortragen. Dann aber will ich das Argument einer „Meta-Meta-Ebene“, auf der Individuen logisch noch vor

³ Im Einklang mit dem Schwinden der Bedeutung des Rationalwahlansatzes bei Jon Elster tritt dieses Problem in letzter Zeit freilich zurück, unter anderem zugunsten Elsters ausführlicher Beschäftigung mit den Emotionen.

⁴ In einem Vortrag am CESifo vor langer Zeit, dessen bibliografische Details ich, shame on me, mittlerweile vergessen habe. Für die Grundhaltung siehe jedoch Binmore (2006).

der Entscheidung über Spielregeln Eigenschaften ihrer Persönlichkeit wählen, als besonderen Beitrag Jon Elsters zur Theorie der Zeitinkonsistenz hervorheben – führt es doch zu dem Schluss, dass die (Fähigkeit zur) Überwindung von Zeitinkonsistenz nicht in jedem Fall wünschenswert sein muss.

Ich beginne mit einer sehr einfachen formalen Analyse von Zeitinkonsistenz (Abschnitt 1) und einer Einordnung dieser in den allgemeinen Kanon von Regelbegründungen und Regelbindungsproblemen (Abschnitt 2). Anschließend setze ich mich mit der Ansicht auseinander, die Überwindung zeitinkonsistenter Versuchungen sei ein Merkmal individueller Rationalität (Abschnitt 3) und werfe mit Blick auf Elsters (1987; 1998) Analyse von Emotionen die Frage auf, ob man bei intrapersonell-intertemporalen Konflikten denn immer Partei für die heutige Inkarnation ergreifen müsse (Abschnitt 4). Eine kurze Zusammenfassung (Abschnitt 5) schließt die Arbeit ab.

1. Eine kurze Theorie des Zeitinkonsistenz-Phänomens

Formalanalytisch betrachtet entsteht das Zeitinkonsistenz-Problem aus einem simplen Grund: Wenn Entscheidungsprobleme nicht blockweise rekursiv sind, fallen die Optimallösungen nach dem Maximumprinzip von Pontryagin und nach Bellmans Gleichung auseinander (Sargent 1987, S. 43 f.). Getrieben wird diese Diskrepanz von *Erwartungen*, denn diese stellen eine inverse Kausalitätsbeziehung in der Zeit her: Erwartetes künftiges Verhalten anderer wirkt auf unser heutiges Tun. Haben diese anderen indes entschieden, so ist ihr Tun versunken, und damit wird mindestens eine Nebenbedingung in unserem Programm aufgeweicht. Es kann dann nicht wundern, dass wir womöglich nachher gerne anders entscheiden (würden), als wir es vorher taten (täten).

Dieses technische Argumentationsmuster hat in der neoklassischen Theorie im Anschluss an die wegweisende Arbeit von Kydland und Prescott (1977) – die unlängst mit dem Nobelpreis für Wirtschaftswissenschaften gekrönt wurden – zahlreiche Anwendungen vor allem im Bereich der Stabilisierungspolitik und der Besteuerung hervorgebracht.⁵

Konstitutionenökonomisch wiederum stellt Zeitinkonsistenz die Kehrseite einer zentralen Begründung für Regeln dar, nämlich der Selbstbindungsvorteile in halbseitigen Dilemmata.⁶ Aus einer wohlverstandenen konstitutionellen Perspektive gehört schließlich zu jeder Begründung für Regeln unauflöslich ein Regelbindungsproblem, denn warum sollte ich künftiges Handeln heute binden wollen, wenn es nicht von meinem jetzt gefassten Plan abweicht? Eine Regel lässt sich folglich überhaupt nicht begründen, ohne in der Begründung selbst die Versuchung für den späteren Regelbruch anzulegen (Beckmann 1998). In einer solchen integrativen Sicht ist die Legitimation für eine Regel untrennbar mit der Frage nach ihrer Stabilisierung verbunden.

⁵ Für einen Überblick und eine vertiefte Diskussion des formalen Hintergrunds siehe Beckmann (1998, Kap. 7), in Grundzügen siehe auch die Diskussion im zweiten Abschnitt.

⁶ Zu deren Charakterisierung siehe Beckmann (2007a) im letztjährigen Band der Reihe „Konzepte der Gesellschaftswissenschaften“ und die Bemerkungen weiter unten.

Spieltheoretisch handelt es sich bei der Zeitinkonsistenz um einen Konflikt zwischen zwei (zeitlichen) Inkarnationen desselben Entscheiders. Diese können mit sich alleine sein – im Falle der *intrapersonellen* Zeitinkonsistenz – oder sich in einer strategischen Aktion mit einem Dritten befinden, dessen Handeln die erste Inkarnation noch beeinflussen will, die zweite indes nicht mehr beeinflussen muss (*interpersonelle* Zeitinkonsistenz). Beide Formen können durch eine – endogene oder exogene – *Veränderung von Präferenzen* entstehen: Das Beispiel des Regierungswechsels aus der Einleitung (Glazer 1989), die Intensivierung des Wunsches nach einem zweiten Glas Burgunder, wenn man das erste genossen hat, aber auch die Zeitinkonsistenz durch nicht exponentielle Diskontierung gehören hierzu.⁷ Neben der Änderung der Präferenzen kommen – bei intrapersoneller Zeitinkonsistenz wohl hauptsächlich – die *Erwartungen* als zentraler Mechanismus in Betracht: Weil man bei der heutigen Planung Rücksicht darauf nehmen muss, wie Erwartungen das Verhalten anderer (oder das eigene Vertrauen, unbesorgt schlafen zu können) beeinflussen, bei der morgigen Durchführung des Plans jedoch nicht mehr, besteht ein Anreiz zur Planrevision.

Nun zu dem formalen Grundmuster einer Spielsituation⁸, in der Zeitinkonsistenz auftreten kann. Betrachten wir ein generisches Zwei-Personen-Spiel mit stetigen eindimensionalen Strategien x^1 und x^2 ,⁹ wobei die Superskripte Individuen bezeichnen. Der Payoff $\pi^i = \pi^i(x^i, x^j)$ des Spielers i hängt offensichtlich sowohl von seiner eigenen Handlung als auch von der seines Gegenübers j ab ($i, j \in \{1,2\}$). In einem Nash-Gleichgewicht mit simultanen Zügen spielen beide Akteure wechselseitig beste Antworten, maximieren also ihren Payoff unter Berücksichtigung der Entscheidung des jeweils anderen. Spieler i maximiert

$$\pi^i = \pi^i(x^i, \bar{x}^j)$$

mit der simplen Bedingung erster Ordnung

$$\frac{\partial \pi^i}{\partial x^i} = 0 \quad (1)$$

In der nachfolgenden Abb. 1 ist diese Lösung grafisch als Punkt A dargestellt. Die gegebene Entscheidung des jeweils anderen Spielers wird durch eine horizontale bzw. vertikale Linie – wie eine „Budgetgerade“ bei gedanklich fixiertem Handeln des Gegenübers – repräsentiert. Dort, wo eine gerade noch erreichbare Indifferenzkurve¹⁰

⁷ Diesem Fall gilt Strotz' (1955) vordringliches Interesse, und er spielt auch bei Elster (1979, S. 36 f.) eine große Rolle.

⁸ Spielsituationen sind Mengen von Spielen, und zwar enthält jede Spielsituation genau diejenigen Spiele, welche (1) identische Spielermengen, (2) identische Mengen *physischer Handlungsmöglichkeiten* (zu unterscheiden von Strategien) und (3) identische Payoff-Funktionen (definiert über Handlungsmöglichkeiten) enthalten. Zwischen ihren Elementen (Spiele) können sich die Regeln – insbesondere die Zugfolgen, Strategien als Kombinationen möglicher Handlungen und die Payoff-Funktionen definiert über Strategiekombinationen – unterscheiden.

⁹ Dies lässt sich ohne Schwierigkeiten auf mehr Dimensionen und eine größere Anzahl von Spielern verallgemeinern, worauf ich hier im Interesse der Lesbarkeit verzichte.

¹⁰ Indifferenzkurven können sich nach oben/rechts öffnen und ein umso höheres Nutzenniveau repräsentieren, je weiter außerhalb sie liegen, oder umgekehrt – wie in der Abb. 1 dargestellt – nach innen geöffnet sein und einen höheren Nutzen verkörpern, wenn sie weiter innen liegen. Letzteres ist beispielsweise im klassischen Cournot-Dyopol der Fall.

diese Linie tangiert, findet sich die beste Antwort auf die gegebene (durch die Linie repräsentierte) Entscheidung des Gegenübers. Die Reaktionsfunktion eines Entscheiders verbindet alle diese seine Optimalpunkte, und im Schnittpunkt der Reaktionsfunktionen haben wir eine Situation wechselseitig bester Antworten, also das Nash-Gleichgewicht bei simultanen Zügen, gefunden. Weil die Indifferenzkurven der Akteure konstruktionsbedingt senkrecht aufeinander stehen, muss diese Lösung bei stetigen Strategien stets Pareto-ineffizient sein (freilich bezogen nur auf die Spieler).

Nehmen wir nun an, ein Spieler (i) könne vor dem anderen (j) ziehen, sodass j die Entscheidung des i vor seinem eigenen Handeln zu beobachten vermag. Für den i ist die Entscheidung des j nicht länger gegeben, sondern abhängig von der eigenen Wahl, und die Bedingung erster Ordnung lautet nunmehr:

$$\frac{\partial \pi^i}{\partial \pi^i} + \frac{\partial \pi^i}{\partial x^j} \cdot \frac{\partial x^j}{\partial x^i} = 0$$

Den letzten Term auf der linken Seite gewinnen wir durch totale Differenziation der Bedingung erster Ordnung für den Stackelberg-Führer j , also

$$\frac{\partial^2 \pi^j}{(\partial x^j)^2} \cdot dx^j + \frac{\partial^2 \pi^j}{\partial x^j \partial x^i} \cdot dx^i = 0$$

Einsetzen führt zu der Bedingung

$$\frac{\partial \pi^i}{\partial x^i} - \frac{\partial \pi^i}{\partial x^j} \cdot \frac{\frac{\partial^2 \pi^j}{(\partial x^j)^2}}{\frac{\partial^2 \pi^j}{\partial x^j \partial x^i}} = 0 \quad (2)$$

In der Abb. 1 stellen wir diese bekannte Lösung als den Punkt B dar. Weil Spieler 1 zuerst zieht und die Entscheidung des ihn beobachtenden 2 in Abhängigkeit von seiner eigenen Wahl rational antizipieren kann, sieht er sich effektiv der Reaktionsfunktion des 2 (und nicht mehr einem jeweils gegebenen x^2) gegenüber und vermag einen beliebigen Punkt darauf auszuwählen. Optimal für ihn ist dann der Punkt, in der seine äußerst erreichbare Indifferenzkurve gerade noEin rascher Blick auf die Abbildung macht deutlich, dass der Payoff des Stackelberg-Führers, der die Initiative ergreift, größer ausfällt als im Punkt A mit simultanen Zügen. Ein zweiter Blick enthüllt sodann, dass, wann immer die Reaktionsfunktion des Folgers 2 eine positive Steigung aufweist, nicht nur der Führer 1 durch seine Selbstbindung gewinnt, sondern auch der Folger 2.¹¹ Unabhängig vom Schicksal des Folgers *lohnt sich die Initiative* genau dann, wenn die Punkte A und B in der Abb. 1 oder anders gewendet die Gleichungen (1) und (2) auseinanderfallen.

¹¹ Das ist der bekannte Fall der strategischen Komplementarität (Gal-Or 1985). In der gängigen Darstellung scheint es oft so, als ob es hier auf die Komplementarität der Instrumente beider (aller) Spieler ankäme. Diese geläufige Darstellung führt indes in die Irre: Dem Führer geht es nämlich immer besser als bei gleichzeitigen Zügen, und allein für die Position des Folgers kommt es darauf an, ob *dessen* Instrument ein strategisches Komplement zu demjenigen des Führers ist.

In einem dritten und letzten Schritt gehen wir nun davon aus, unser Entscheider 1 habe erfolgreich die Initiative ergriffen und den Punkt B in der Abb. 1 erreicht. Allerdings habe er nun *für den Folger 2 überraschend* die Möglichkeit, seine Entscheidung zu überdenken. Wenn dem so wäre, erschiene ihm die optimale Entscheidung des Folgers 2 nunmehr als *gegeben*. Die Bedingung erster Ordnung für das neue Problem entspräche wiederum der Gleichung (1) – wengleich bei einem anderen Niveau von x^2 –, und der Stackelberg-Führer würde in dieser Lösung (Punkt C in der Abb. 1) eine noch bessere Lösung erreichen, nun allerdings eindeutig zu Lasten des Folgers 2 (verglichen zum Punkt B). *Diese Versuchung „nachzukarten“ stellt den Kern des Zeitinkonsistenz-Problems dar.*

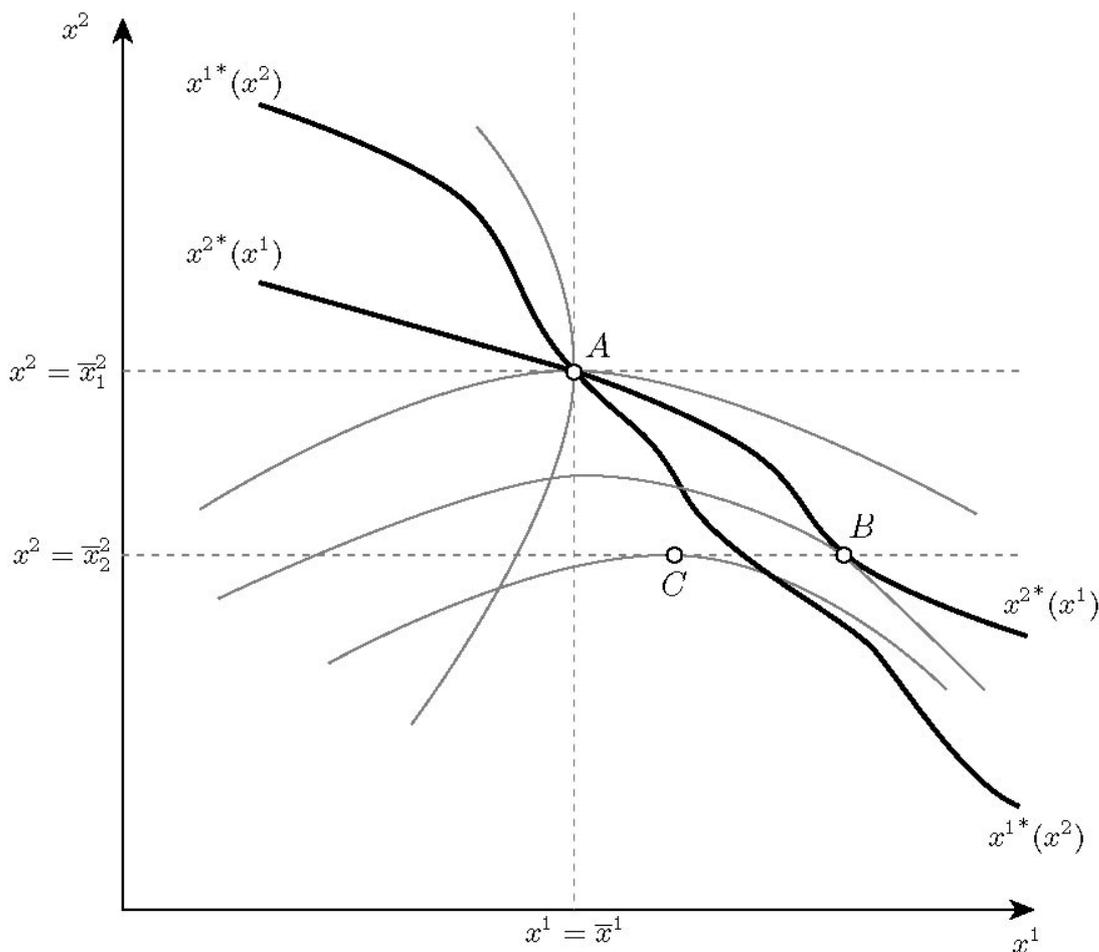


Abbildung 1: Simultane vs. sequenzielle Züge in einem generischen Spiel

Wenn wir uns diesem Problem wie dargestellt annähern, werden zwei wesentliche Punkte deutlich:

- Zeitinkonsistenz stellt das Spiegelbild der Vorteile von Selbstbindung dar. Wann immer es sich lohnt, sich vorab auf eine spätere Handlung festzulegen, lohnt es sich auch, diese Entscheidung später zu revidieren. Das ist der Kern der eingangs zitierten Bemerkung von Strotz (1955).
- Damit haben wir es nicht mit einem pathologischen Sonderfall zu tun, sondern mit einer allgemeinen Eigenschaft von Regeln, die Selbstbindungsvorteile nutzen, und einer allgemeinen Eigenschaft von Stackelberg-Führerschaft.

Um einem Missverständnis gleich zu Anfang vorzubeugen: Die Geschichte, die ich hier erzähle, läuft nicht darauf hinaus, dass ein rationaler Entscheider *tatsächlich* ausgehend von Punkt A die Initiative ergreift, um den Punkt B zu erreichen, sodann aber seinen rationalen Gegenspieler düpirt und aus der Nachhand Punkt C realisiert. *In der Praxis* mag dies vorkommen, indem ein Spieler das Vertrauen¹² anderer in seine Ankündigungen enttäuscht, *spieltheoretisch gesehen* bestehen indes nur zwei Möglichkeiten (in reinen Strategien): Entweder es gelingt einem Spieler, eine nachträgliche Abweichung von seiner Wahl unmöglich zu machen (wirksames Commitment), und wir gelangen zu der Lösung B, oder dies gelingt nicht, und es bleibt beim Nash-Gleichgewicht in simultanen Zügen A.

Mit dieser allgemeinen spieltheoretischen Analyse sind wir auch in der Lage anzugeben, unter welchen Ausnahmbedingungen kein Zeitinkonsistenz-Problem auftritt. Dazu bedarf es nur eines Vergleichs der einfachen Bedingungen (1) und (2) – ist der zweite Term auf der linken Seite von (2) gleich Null, so fallen die Bedingungen (wie auch die Punkte A, B und C in der Abb. 1) zusammen.

Trivial ist zunächst der Fall: $\partial \pi^i / \partial x^j = 0$ Besteht keine Reaktionsverbundenheit zwischen den Akteuren – genauer: hat die Entscheidung des j also keinen externen Effekt auf den i –, kann Commitment sich auch nicht lohnen. In einer strategischen Situation dagegen bestehen drei weitere Möglichkeiten:¹³

- *Teamlösungen*. Sind die Präferenzen der Akteure identisch, so führt die Maximierung des Nutzens durch den j zur Bedingung $\partial \pi^j / \partial x^j = \partial \pi^i / \partial x^j = 0$
- *Dominante Strategie des Folgers*. Falls ein Spieler eine dominante Strategie hat, könnte ein Stackelberg-Führer keinen indirekten Einfluss auf sein Verhalten ausüben, indem er seine eigene Entscheidung variiert. Dies entzieht einer Commitment-Strategie die Grundlage. In der Abb. 1 verlief die Reaktionsfunktion des Individuums 2 horizontal, und formal hätten wir $\partial^2 \pi^j / (\partial x^j \partial x^i) = \infty$.
- *Punkt A ist der bliss point des Stackelberg-Führers*. Ist bei simultanen Zügen zufällig die allerbeste Lösung für einen Spieler erreicht – sodass jede Bewegung aus A ihn auf ein niedrigeres Indifferenzniveau brächte¹⁴ –, so besteht ebenfalls kein Grund mehr zur indirekten Einflussnahme auf das Gegenüber.

In allen übrigen Fällen – und damit in den theoretisch interessantesten Situationen – lohnt sich Commitment, und damit besteht *uno actu* ein Zeitkonsistenz-Problem.

¹² Siehe Lahno (2002) zu diesem Begriff.

¹³ Für eine ausführliche Diskussion vgl. Beckmann (1998, S. 442–452).

¹⁴ In der Abb. 1 wäre dies offenkundig dann der Fall, wenn die Indifferenzkurven des betreffenden Individuums Konturen um den Punkt A herum bildeten. Auch dann existierte eine Reaktionsfunktion, gebildet nach dem beschriebenen Schema, doch wäre eine Bewegung entlang ihres Grafen dann nicht mehr mit monoton steigenden oder fallenden Payoffs verbunden.

2. Regelbegründungen und Regelbindungsprobleme

Es könnte so scheinen, als handele es sich bei dem engen Zusammenhang zwischen Regelbegründung und Regelbindungsproblem um ein Spezifikum von Selbstbindungen. Indes lässt sich der Grundgedanke auf alle anderen Regelbegründungen übertragen, wie ich in diesem Abschnitt verdeutlichen will. Die Tab. 1 – siehe auch Beckmann (2007a) zeigt dazu zunächst eine *distinctio completa* der Fälle, in denen ein zumindest begrenzt rationaler Entscheider einer Bindung seines eigenen künftigen Verhaltens zustimmen kann – also auf einer Regelebene bestimmte physisch mögliche Optionen für die nachfolgende Handlungsebene ausschließen wollen kann. Wir unterscheiden zunächst danach, ob bei der betrachteten Regel die Gestaltung von Anreizen oder aber das Wissensproblem (die Sammlung, Verarbeitung und Übertragung von Informationen) im Mittelpunkt steht. Eine zweite Dimension erhält man, wenn man zwischen Regeln unterscheidet, welche sich der Entscheider auch außerhalb eines gesellschaftlichen Kontextes auferlegen würde, und solchen, die nur im Rahmen einer gesellschaftlichen Interaktion sinnvoll erscheinen.

Tabelle. 1: Systematik der Begründungen von Regeln

	Individuell	Kollektiv
Anreize	Halbdilemmata: die eigene Bindung beeinflusst andere indirekt (instrumentell), wobei „andere“ auch die eigene Person zu einem späteren Zeitpunkt umfasst	
		Lösung von Gefangenen-Dilemma-(PD-)Situationen
		Schaffen von PD-Situationen (Pies, 2003)
Wissen	Verbesserung der Entscheidungsfindung durch Reduktion von Komplexität (Levy, 1992)	Speicherung und Gebrauch impliziten Wissens; Förderung von Experimenten und evolutionären Prozessen (Hayek, 1991)
	Steigerung oder (gewünschte) Senkung der Vorhersehbarkeit (Heiner, 1983), Reduktion oder (gewünschte) Steigerung der Variabilität	

Im Ergebnis kommt man zu vier wesentlichen normativen Begründungen für Regeln:

- Regeln können als *Daumenregeln* zur Verbesserung von Entscheidungen beitragen, wenn die Individuen nur begrenzt statt vollständig rational sind.
- Sie können darüber hinaus als Wissensspeicher dienen und die *Nutzung impliziten Wissens* ermöglichen: Sofern eine Regel gesellschaftliches Wissen verkörpert, beispielsweise durch langsame Fortbildung im Rahmen der gesellschaftlichen Evolution,¹⁵ setzt dieses Wissen auch derjenige dieses Wissen ein, der sich rein mechanisch und unreflektiert an die Regel hält.¹⁶
- Weiters erhöhen Regeln die *Vorhersehbarkeit* des Verhaltens anderer (Heiner 1983). Anders als bei den beiden vorgenannten Mechanismen geht es hier nicht um die Verbesserung der Entscheidungen eines Robinson Crusoe, sondern zum ersten Mal auch um soziale Interaktion – gleichwohl steht immer noch die Erleichterung des individuellen Kalküls im Vordergrund. Eine verlässliche Bindung anderer an Regeln strukturiert und vereinfacht das spieltheoretische Optimierungskalkül eines Individuums, dessen persönlicher Nutzen auch vom Verhalten der anderen abhängt.
- Die vierte Kategorie von Mechanismen verweist uns vollends auf den Bereich sozialer Interaktion. Geht es hier doch um die *Schaffung von Anreizen* für andere, sich auf eine (individuell oder sozial) wünschenswerte Weise zu verhalten. Es lassen sich hier zwei Untergruppen unterscheiden: Bei der ersten, den so genannten *Halbdilemmata*, dient die glaubwürdige Bindung des eigenen Verhaltens als Mittel, das Verhalten anderer *indirekt* zu beeinflussen. Im Falle *sozialer Dilemmata* dagegen bezweckt die Unterwerfung unter die Regel die *direkte* Beeinflussung des Verhaltens anderer, und die Bindung des eigenen Verhaltens stellt den Preis dar, welcher dafür zu zahlen ist.

Offensichtlich beziehen sich unsere Überlegungen aus Abschnitt 1 zu Selbstbindung und Zeitinkonsistenz auf den Fall der Halbdilemmata. Um dies zu verdeutlichen, zeige ich in der nachfolgenden Tab. 2 einen der beiden paradigmatischen Fälle (anders als

¹⁵ Auch hinsichtlich der Einschätzung gesellschaftlicher Evolution lohnt es sich, Elster zu lesen. Meine eigene Skepsis hinsichtlich Hayek'scher und anderer Varianten einer evolutorischen Ordnungsökonomik speiste sich bislang vor allem aus den Schwierigkeiten, Konzepte der Schumpeter'schen Evolutionsökonomik (der Märkte) auf Ordnungen zu übertragen (Beckmann 2007b, S.74–82): Zweifel kamen mir angesichts der Notwendigkeit von Gruppenselektion einerseits und dem Fehlen eines (Ordnungs-)Rahmens für den Systemwettbewerb andererseits. Elster (2007, S. 271–298) äußert weitere, tiefer sitzende Bedenken: Danach sei die natürliche Evolution nicht imstande, lokale Maxima zugunsten globaler Maxima zu „überspringen“, und Mutation und Selektion bedürften langer Zeit, die anders als bei der natürlichen Evolution im gesellschaftlichen Kontext nicht zur Verfügung stünde. Diese Kritik erfasst nicht nur die Übertragung Schumpeter'scher Argumentation auf den Systemwettbewerb, sondern jede Form von sozialer Evolution, auch Schumpeters (2005) Ansatz selbst. (Allerdings muss ich auch Elster (2007) widersprechen: Rationale Entscheidungen von Individuen allein sind nicht hinreichend für das Erreichen eines globalen Maximums – dazu bedarf es neben der Rationalität noch voller Information. Siehe die Argumentation zu den Daumenregeln im Text.)

¹⁶ Mein Lieblingsbeispiel ist die Benimmregel, Kartoffeln nicht mit dem Messer zu zerteilen. Diese Regel dürfte sich aufgrund des Umstands entwickelt haben, dass Silbermesser aufgrund des hohen Stärkegehaltes von Kartoffeln anlaufen bzw. (bei Eisenmessern) rosten, wenn man sie zum Kartoffelschneiden benutzt. Wer sich nur mechanisch aufgrund seiner „guten Kinderstube“ an diese Regel hält, gebraucht damit immerhin einiges Wissen über physikalisch-chemische Zusammenhänge, ohne sich dieser überhaupt nur bewusst sein zu müssen. Freilich verdeutlicht dieses Beispiel auch, dass Regeln als Wissensspeicher überleben können: Für Messer aus rostfreiem Stahl macht die Regel keinen Sinn mehr.

zuvor in der Abb. 1 nützt hier die Selbstbindung des Stackelberg-Führers – des Spaltenspielers – auch dem Folger). Einziges Nash-Gleichgewicht in diesem Spiel ist die orange hinterlegte Lösung { unten ; rechts }, dies entspricht dem Punkt A in der Abb. 1. In diesem Gleichgewicht bleibt freilich der Spaltenspieler auf sein zweitschlechtestes Ergebnis verwiesen.

Tabelle. 2: Halbdilemma in ordinaler Darstellung (Fall des Pareto-superioren Commitment)

		Spaltenspieler	
		links ↓	rechts
Zeilenspieler	oben	(4 , 3)	(1 , 4)
	unten	(2 , 1)	(3 , 2)

Ursächlich für dieses Problem des Spaltenspielers ist, dass er *zu seinem eigenen Nachteil* über eine dominante Strategie – „rechts“ – verfügt: zu seinem eigenen Nachteil deshalb, weil der Zeilenspieler angesichts der klar erkennbaren Anreize des Spaltenspielers rational die Option „unten“ wählen muss, um nicht seinerseits auf den Minimal-Payoff verwiesen zu werden. Könnte sich der Spaltenspieler glaubhaft auf „links“ festlegen – angedeutet durch einen Pfeil in der Tab. 2 –, so würde {oben; links} realisiert (dem Punkt B entsprechend), was in diesem Fall¹⁷ für beide Akteure besser wäre als als die Lösung mit gleichzeitigen Zügen. Allerdings verbliebe für den Spaltenspieler die Versuchung, einseitig auf „rechts“ zu wechseln und damit seine bestmögliche Lösung (dem Punkt C in der Abb. 1 entsprechend) zu realisieren.

Dass der letzte Schritt das Regelbindungsproblem einer Selbstbindung darstellt, sollte mittlerweile hinreichend klar sein. Ebenso bleibt es dem geneigten Leser überlassen, die Verbindung zum klassischen Gefangenen-Dilemma zu knüpfen, in dem beide Akteure über eine selbst schädigende dominante Strategie verfügen. Das offensichtliche Regelbindungsproblem besteht in diesem Fall in der altbekannten Versuchung zum *free riding*.

In der Literatur wurde dieser Zusammenhang gelegentlich erkannt (siehe auch Beckmann 1998, Kap. 5). Elster (2000, S. 40) selbst spricht im Zusammenhang mit der Zeitinkonsistenz von *dynamischen Gefangenen-Dilemmata*. Das erkennt allerdings, dass es sich, wie oben gezeigt, um eine systematisch verschiedene Spielsituation handelt. Unter anderem muss Commitment nicht zu einer Pareto-superioren Lösung führen (siehe Abb. 1). Genauer erscheint die Unterscheidung zwischen Gefangenen-Dilemmata und „Quasi-PDs“ bei Pies (2000) (op.cit., S. 222 f.), wobei letztere

¹⁷ Mindestens das Instrument des Zeilenspielers ist also ein strategisches Komplement zu dem des Spaltenspielers.

weitgehend identisch sind mit dem, was wir hier als Halbdilemma bezeichnen. Daneben führt Pies (2000, S. 43 ff.) noch eine Kategorie in extensiver Form ein: das „einseitige“ Gefangenen-Dilemma. Nun ist freilich das in diesem Zusammenhang dargestellte Markteintrittsspiel (Binmore 2006, S. 171 f.) letztlich eine bestimmte Konkretisierung – ein Element – der Spielsituation aus der Tab. 2 mit der kleinen Änderung, dass beide Spieler zwischen { unten ; links } und { unten ; rechts } indifferent sind.¹⁸ Eine eigene Kategorie scheint mir damit nicht vorzuliegen, schon gar nicht im Sinne einer Gruppe von Spielsituationen.

Etwas schwerer als bei den anreizbasierten fällt es bei den wissensbasierten Regelbegründungen, den unmittelbaren Zusammenhang zwischen Vorteilhaftigkeit der Regel und Verlockung zum Regelbruch zu zeigen. In der Tab. 3 werden die verschiedenen Kategorien entsprechend der Systematik der Tab. 1 dargestellt.¹⁹

Tabelle. 3: Systematik der Regelbindungsprobleme

	Individuell	Kollektiv
Anreize	Zeitinkonsistenz	
		free riding
Wissen	Desorientierung?	Vernichtung von Sozialkapital
	Vertrauensverlust	

Etwas aus der Reihe fällt das Regelbindungsproblem bei Daumenregeln, welches überdies mit einem Fragezeichen versehen ist. Hier dürfte sich eine kurze Erläuterung lohnen, zumal dies auch einen Blick auf das Wesen begrenzter Rationalität eröffnet und eine Beziehung zu einem Argument Elsters (2007) gegen evolutorische Ansätze in den Sozialwissenschaften besteht (siehe die Fußnote 11).

Betrachten wir einen Entscheider, der seinen Netto-„Lohn“ durch den Einsatz von „Mühe“ zu maximieren trachtet (Abb. 2).²⁰ Der Zusammenhang zwischen Lohn und Mühe sei durch die Zielfunktion $z(f(\bullet))$ gegeben. Diese allerdings ist nicht konkav, sodass neben dem globalen Maximum B noch ein zweites lokales Maximum A existiert.

¹⁸ Der Grund hierfür liegt darin, dass das Spiel abbricht, nachdem der Zeilenspieler „unten“ gezogen hat.

¹⁹ Leser von Beckmann (1998) mögen sich fragen, wo die Schleivorteile bzw. Schleierprobleme, welche durch die Möglichkeit der Risikokonsolidierung bei unsicheren Erträgen i.w.S. entstehen, geblieben sind. Bei näherem Hinsehen handelt es sich hier indes um eine spezielle Erscheinungsform der Kooperation in sozialen Dilemmata mit dem Unterschied, dass die zu wählende kooperative Strategie nun zustandsabhängig ist. Für die Darstellung in diesem Aufsatz bedarf es keiner weiteren Unterkategorie.

²⁰ Vgl. dazu Levy (1992).

Für einen rationalen Entscheider mit vollständiger Information stellt die Nicht-Konvexität des Möglichkeitsraums kein Problem dar – er wird nicht nur auf die Marginalbedingungen erster Ordnung schauen, sondern darüber hinaus auch die Totalbedingungen berücksichtigen (Elster 2007).²¹ In vielen Fällen wird der Entscheider jedoch gerade nicht über vollständige Informationen verfügen – gängig erscheint vielmehr der Fall *lokaler Information*, in dem der Entscheider das Verhalten des Systems in der Umgebung des Status quo kennt, aber nicht das Ergebnis größerer Strategieänderungen oder Reformen. Dann verbleibt ihm aber nur noch die Möglichkeit, sich an den Bedingungen erster Ordnung zu orientieren, getreu der alten Devise, dass ein Bergsteiger im Nebel dann noch nicht auf dem Gipfel angekommen ist, wenn er durch einen Schritt in irgendeine Richtung weiter aufsteigen kann.

Durch diese Informationsbeschränkung gelangen wir zu einer Fallgruppe der *begrenzten Rationalität*. Kennzeichnend dafür ist ein Auseinanderklaffen der Anforderungen des Problems – bedingt vor allem durch dessen Komplexität – und der Optimierungs- und Entscheidungsfähigkeiten des Entscheiders.²² Im Ergebnis kommt es nun auf den Ausgangspunkt an, ob das rein lokale Maximum A oder das globale Maximum B erreicht wird. Setzen wir diese Situation voraus, so könnte die Einführung einer *Daumenregel* dem Entscheider helfen, seine Wahl zu verbessern, wie dies in der Abb. 2 gezeigt ist. Nehmen wir an, unser Entscheider habe eine religiöse Vorschrift verinnerlicht, nach der der „gerechte“ Lohn, den man höchstens fordern dürfe, ohne zu sündigen, eine lineare Funktion $cc(\bullet)$ der aufgewendeten Anstrengung sei. Damit wird nun der Punkt C erreicht, der zwar kein globales Maximum des ursprünglichen Problems darstellt, aber immerhin zu einer besseren Lösung führt, als im lokalen Maximum A realisiert wird. Startete der Entscheider im Ursprung, wäre er gut beraten, die Regel zu akzeptieren. Wir erkennen auch leicht, wie diese Regel funktioniert: Das Programm

$$\begin{aligned} \max l &= z(m) \\ \text{s.t. } l &\leq cc(m) \end{aligned} \quad (3)$$

ist konvex – eine methodische Vereinfachung, die wir mit Hilfe der Ergänzung des ursprünglichen Optimierungsproblems um die Nebenbedingung cc erreicht haben.

²¹ Vgl. dazu auch Elster (1987, S. 33–65).

²² Die „Kompetenz-Schwierigkeits-Lücke“ (Heiner 1990). Die „Kompetenz“ hängt ihrerseits von den kognitiven Fähigkeiten des betrachteten Individuums, einschließlich bestimmter systematischer Verzerrungen, und andererseits von den verfügbaren Informationen ab (Beckmann 1998). Somit kann auch ein Individuum mit göttlichen Fähigkeiten zur Verarbeitung von Informationen und übermenschlicher Selbstbeherrschung begrenzt rationale Entscheidungen fällen.

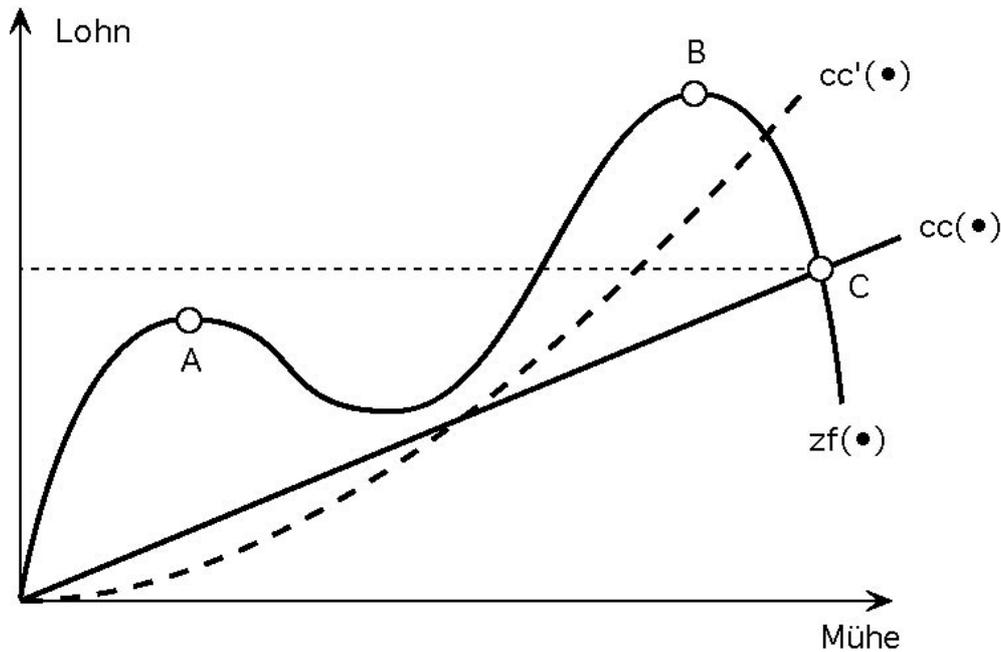


Abbildung 2: Daumenregel und Regelbindungsproblem

So weit zur Regelbegründung. Aber wie steht es mit dem zugehörigen Regelbindungsproblem? Zunächst ist klar: Die Punkte B und C fallen nur im Ausnahmefall zusammen, im Allgemeinen wird also auch hier die Versuchung bestehen, *ex post* die Regel zu brechen. (Dazu genügt auch lokale Information!) Fraglich bleibt, worin genau ein *Problem* bestehen soll. Ein solches entstünde freilich stets dann, wenn man im Bewusstsein des späteren „Nachsteuerns“ keine Bindungswirkung der ursprünglichen Daumenregel erreichen könnte, wenn also ein *zweistufiges Entscheidungsverfahren*, in dem die Anwendung einer Daumenregel im ersten Schritt mit freier Optimierung in einem zweiten Schritt verbunden wird, aus letztlich psychologischen Gründen nicht zu Gebote steht. Ebenfalls angedeutet ist in der Abb. 2 eine mögliche Weiterentwicklung der Daumenregel (hin zu $cc'(\bullet)$), welche durch Nutzung der gesammelten *Erfahrungen* zustande kommen könnte.

Diese Fragen auszuloten trüge allerdings nur mehr wenig zu einer Würdigung des Beitrags Jon Elsters zur Theorie der Zeitinkonsistenz bei. Im Augenblick kommt es mir neben der Einordnung in die Systematik der Regelbindungsprobleme zuerst auf einen Einstieg in die Frage der begrenzten Rationalität an. Denn hier liegt der Schlüssel zu Elsters Sichtweise des Problems.

3. War der listenreiche Odysseus wirklich nur begrenzt rational?

Was wir grundsätzlich unter „Rationalität“ verstehen, erscheint auf den ersten Blick klar. Ein Individuum kann als rational gelten, wenn es unter bewusstem Einsatz seines Verstandes bemüht ist, unter den verfügbaren Optionen die bestmögliche auszuwählen, einschließlich einer endogenen Beschaffung und Verarbeitung von Informationen. Auch der mikroökonomische Rationalitätsbegriff kann im Grundsatz als einfach gelten: Konstitutiv für Rationalität sind Vollständigkeit, Reflexivität und Transitivität von Präferenzen, da sie dann (bei endlichen Möglichkeitsräumen) durch eine geeignet

konstruierte Nutzenfunktion repräsentiert werden können. Bei näherem Hinsehen jedoch löst sich das schöne Bild auf wie ein pointilistisches Gemälde – von Paradoxien bei Entscheidungen unter Unsicherheit über Entscheidungsheuristiken bis hin zu der hier relevanten Frage nach der Transitivität in der Zeit tauchen problematische Punkte auf. Über daraus resultierende Konzepte von begrenzter Rationalität existiert mittlerweile eine unübersehbare Literatur. In Elsters Auseinandersetzung mit der Zeitinkonsistenz steht jedoch *ein* Aspekt im Vordergrund:

Wie wir gesehen haben, birgt jeder Vorteil, der eigenen Bindung an eine Regel zuzustimmen, bereits die Versuchung zum Regelbruch in sich. Im Fall von Selbstbindungen bietet die Möglichkeit, diese Versuchung zu überwinden, den Schlüssel zur Glaubwürdigkeit. Nach allgemeinem Dafürhalten gibt es grundsätzlich zwei Wege, dies zu erreichen: Erstens kann man sich durch *äußere Hilfsmittel* der Möglichkeit zum Regelbruch berauben, zweitens kann man die *innere Stärke* aufbringen, der Versuchung zum Regelbruch zu widerstehen. Elsters Meinung nach lässt sich nur letzteres als vollständige Rationalität bezeichnen. Nirgends formuliert er dies deutlicher als in Elster (1987, S. 67–68)²³:

Odysseus war nicht gänzlich rational, denn ein rationales Wesen hätte nicht zu diesem Mittel greifen müssen; er war aber auch nicht einfach der passive und irrationale Spielball seiner wechselnden Wünsche und Bedürfnisse, denn er konnte mit indirekten Mitteln das gleiche Ziel erreichen, das eine rationale Person auf direkte Weise realisiert hätte. Sein Dilemma – schwach zu sein und davon zu wissen – verweist auf die Notwendigkeit einer Theorie *unvollständiger Rationalität* [...].

Ich vertrete folgende Hauptthese: Selbstbindung ist ein vorrangiger Weg, das Problem der Willensschwäche zu lösen, die entscheidende Technik, Rationalität durch indirekte Mittel zu erlangen. Es gibt jedoch noch einen anderen Weg, den man einschlagen kann. Diese Alternative besteht, kurz gesagt, in einer Umstrukturierung des inneren Raums der Person, ohne dass irgendwelche kausalen Mechanismen in der äußeren Welt errichtet werden. [...] Ich möchte nicht bestreiten, dass man ein gewisses Maß an Selbstbeherrschung einfach dadurch erreichen kann, dass man sich am Riemen reisst, aber ich werde zeigen, dass sich dauerhaftere Ergebnisse erreichen lassen, wenn man die Umwelt in sein Handeln mit einbezieht.

I beg to differ. Natürlich haben wir es ein Stück weit bloß mit einer Definitionsfrage zu tun, und wie der Dichter sagt: „Mit Worten lässt sich trefflich streiten, mit Worten ein System bereiten, an Worte lässt sich trefflich glauben, von einem Wort kann man kein Jota rauben.“ Grundsätzlich könnten wir die Bezeichnung „Rationalität“ gerne an die Erfüllung übermenschlicher Anforderung knüpfen und real existierende Vernunft unter der Rubrik „begrenzte Rationalität“ erfassen. Alles, was nicht bewusst unter Einsatz des Verstandes nach einer möglichst guten Lösung sucht, hieße dann „nicht rational“.

Auf dieser rein semantischen Ebene kann ich nur zwei Bedenken anführen: Erstens erweitert *uno actu* den Bereich begrenzter Rationalität, der den der vollständigen Rationalität einschränkt. Damit muss die erstgenannte Kategorie mehr Phänomene aufnehmen, die sich untereinander möglicherweise stark unterscheiden. Zweitens mag man zwar rationale Entscheidungen vor allem als ein normatives Konzept verstanden haben, doch ist daneben auch die Vorstellung des Rationalitätskonzepts als einer *positiven Theorie* menschlichen Verhaltens *im Durchschnitt* vertreten (Elster 1986, S. 2

²³ Hervorhebung im Original. Bei Elster (1987) handelt es sich um eine Kompilation deutsch übersetzter Texte aus *Sour Grapes* (Original: Elster 1983) und *Ulysses and the Sirens* (Original: Elster 1979). Meine erste Berührung mit Elster kam zustande, als ich während meines Studiums genau dieses Buch in der Bibliothek meiner Eltern fand.

f.). Dies würde gleichfalls erschwert, stellte man auf übermenschliche Willenskraft als Definitionsmerkmal von Rationalität ab.

„Doch ein Begriff muss bei dem Worte sein!“²⁴ Aus der theoretischen Betrachtung von Regelbegründungen und Regelbindungsproblemen lassen sich zwei tiefer greifende Argumente gegen Elsters oben zitierte Position²⁵ ableiten:

- Wer die Selbstüberwindung bei *individueller* Zeitinkonsistenz – also im Spiel eines Entscheiders gegen sich selbst, um das es Elster vordringlich zu gehen scheint – als Merkmal rationalen Verhaltens fordert, könnte aufgrund der formal gleichen Struktur die *Wahl einer dominierten Strategie* durch einen rationalen Akteur zumindest für Halbdilemmata, womöglich aber auch für PD-Situationen fordern. Damit gäbe es aber kein soziales Problem mehr, das rationale Entscheider nicht ohne jeden Rückgriff auf gesellschaftliche Institutionen dadurch lösen könnten, indem sie einfach i.w.S. kooperative Strategien auch bei Bestehen adverser Anreize wählen.
- Nehmen wir einmal an, ein rationales Individuum habe *auf der Meta-Ebene* die Wahl, ob es die psychologischen Eigenschaften annehmen wolle, welche die Überwindung von Zeitinkonsistenz erleichtern. Es könne also nicht nur über Handlungen und (sekundäre) Verhaltensregeln, sondern auch über die „fest verdrahteten“ (primären) Regeln entscheiden.²⁶ Bei diesem Entschluss muss sich das Individuum nicht notwendig auf Elsters (1987) „Umstrukturierung des inneren Raums“ festlegen – und zwar deshalb nicht, weil diese Persönlichkeitsveränderung mit einer Herabsetzung der Genussfähigkeit verbunden sein kann.

Das zweite Argument wird im vierten Abschnitt behandelt. Wenden wir uns zunächst dem ersten Punkt zu! Grundlage für diese Betrachtung ist der Umstand, dass die *formale Struktur unverändert* bleibt. Es ist gleich, ob wir eine Interaktion zwischen zwei Individuen – Stackelberg-Führer und Stackelberg-Folger – oder zwei Inkarnationen desselben Individuums betrachten. Für die typische Anwendung auf einen Raucher oder einen übergewichtigen Menschen, welcher der Versuchung zu „sündigen“ widerstehen muss,²⁷ tritt in der Normalform aus der Tab. 2 (auf der Seite 10) die heutige, langfristig planende Inkarnation (genannt „Planer“) in die Rolle des Zeilenspielers, die spätere, der Versuchung ausgesetzte Inkarnation (genannt „Sünder“) in die Rolle des Spaltenspielers. Die Strategie „links“ entspricht offensichtlich dem Verzicht, die Wahl „rechts“ dem Nachgeben gegenüber der Versuchung – eine dominante Wahl, weil im Fall geplanter Abstinenz „einmal keinmal“, im Fall eines

²⁴ Gerade dieses Zitat aus Goethes Faust darf man nicht ohne die Antwort des Mephistopheles belassen: „Schon recht, doch muss man sich nicht ängstlich quälen, denn gerade wo Begriffe fehlen, da stellt ein Wort zur rechten Zeit sich ein.“

²⁵ Und vergleichbare Positionen anderer Autoren, etwa McClennan (1990) und Gauthier (1985).

²⁶ Dies ist weniger kontrafaktisch, als es auf den ersten Blick erscheint. Ein Stück weit beeinflussen wir durch unser Verhalten, besonders durch langfristige Wahlentscheidungen, wie zum Beispiel die Auswahl einer Laufbahn, auch die Ausformung und Entwicklung unserer Persönlichkeit. (Siehe auch Brennan und Buchanan 1993. Harrison 2006 erklärt Selbstmordanschläge aus dem Versuch hoffnungsloser junger Menschen, eine Persönlichkeit und eine bedeutsame Rolle für sich zu erlangen. Dem hat jüngst Matteucci 2007 mit typischer Ökonomie-Kritik widersprochen.)

²⁷ In der Literatur diskutiert man eine Vielzahl von Beispielen. Die beiden im Text habe ich gewählt, weil sie Elster (1987, S. 69) anführt. Man beachte, dass die Abb. 2 eine Situation widerspiegelt, in dem die Selbstbindung für beide Inkarnationen des betrachteten Individuums besser ist als der Verzicht auf Selbstbindung. Das entspricht der materiellen Geschichte Elsters (op.cit.).

Verzichts auf Abstinenz „eh alles egal“ ist. Entsprechend verfügt der Planer über die Strategien „Fang ein neues Leben an“(oben) und „Status quo“ (unten).

Die Normalform in der Tab. 2 definiert zwar die *Spielsituation*, aber nicht die dynamische Struktur der Interaktion. Punkt B in der Abb. 1, in dem es zu Commitment kommt, läuft darauf hinaus, dass *der Sünder zuerst zieht* und sich seiner dominanten Strategie beraubt. In der Praxis wird dies freilich nicht der Sünder – den es ja realiter noch nicht gibt –, sondern der Planer für den Sünder tun, der sich damit zugleich die Voraussetzungen für eine Wahl der Abstinenz schafft (wie das Stellen eines Weckers und die Unmöglichkeit, diesen später im Halbschlaf abzuschalten, die Voraussetzung für einen ruhigen Schlaf darstellt). Dies macht klar, dass wir eine Form *prozeduraler Rationalität* diskutieren, in der nicht mehr einzelne Handlungen, sondern komplette Pläne Gegenstand rationaler Wahl darstellen; ein prozeduraler Entscheider wählt denjenigen Plan *und führt ihn durch*, der im Barwert (mit oder ohne Nutzendiskontierung!) den höchsten Payoff bietet.²⁸

Für eine Diskussion sollten wir uns zuerst vergegenwärtigen, dass diese strategische Formulierung eine Reihe von Situationen abbilden kann, die gleichwohl materiell nicht völlig äquivalent sind:

- In Newcombs Paradox hat der Zeilenspieler schon gezogen – der Kniff besteht hier in der Annahme seiner übernatürlichen Allwissenheit, wodurch die Wahlhandlung des Spaltenspielers *selbstbezüglich* wird.²⁹ Dass solche

²⁸ Eine Erscheinungsform der Regelnehmerschaft (Beckmann 1998), die sich zu einer handlungsbezogenen Rationalität etwa so verhält wie der Regelparetianismus (e.g. Harsányi 1985) zum herkömmlichen Pareto-Kriterium.

²⁹ Siehe zum Setup Binmore (1994, S. 252): Der geneigte Leser möge annehmen, ihm erscheine unverhofft ein höheres Wesen, das sein Verhalten mit völliger Sicherheit vorhersagen könne. Das Wesen zeige zwei verschlossene Holzkästchen vor, die durch je ein Schild als „A“ bzw. „B“ bezeichnet seien, und biete dem Leser an, entweder Kästchen A oder beide Kästchen zu nehmen. In Kästchen B befinde sich auf jeden Fall ein 500-Euro-Schein. Ferner habe es im Vorhinein die Absichten des Lesers erforscht und in das Kästchen A einen Diamanten im Wert von einer Million Euro gelegt, falls der Leser sich nur für Kästchen A entscheiden werde, ansonsten bleibe Kästchen A leer. Was soll der Leser tun? Das Dominanzprinzip empfiehlt ihm, beide Kästchen zu nehmen. Denn so oder so hat das höhere Wesen die Kästchen bereits in der Vergangenheit bestückt, und die Kästchen enthalten jetzt zusammengenommen entweder Werte in Höhe von 1.000.500 oder 500 Euro. Welcher Fall auch immer vorliegen mag, der Leser stellt sich um 500 Euro besser, wenn er beide Kästchen nimmt. Der offensichtliche Haken ist, dass das höhere Wesen dies antizipiert haben wird und sich der Diamant deswegen nicht in Kästchen A befindet. Dass dieses Beispiel die formale Struktur aus dem Abschnitt I hat, ist offensichtlich. Wie gehabt stellt sich der rationale Entscheider, der dem Dominanzprinzip anhängt, gewissermaßen selbst ein Bein; und wie dort ist es die Transparenz der kognitiven Prozesse des Entscheiders, die das Problem erst entstehen lässt.

Im Zusammenhang mit diesem Paradox entstehen interessante Fragen: Inwieweit kann man in der hypothetischen Situation, in die ich den Leser gestellt habe, überhaupt davon ausgehen, dass dieser noch einen freien Willen (freie Hand bei der Entscheidung) habe, obgleich doch die ex hypothesi korrekte Prognose des höheren Wesens bereits vorliegt? Es ist ja gar nicht möglich, dass der Entscheider von dieser gegebenen Prognose abweicht. Und welche Vorstellung von Kausalität legt man zugrunde, wenn man zukünftige Taten des Lesers auf vergangene Taten des höheren Wesens rückwirken lässt? Es ist schwer vorstellbar, dass der Entscheidung des Lesers zum einen eine gegebene Bestückung der Kästchen zugrunde liegt, zum anderen aber gerade die Entscheidung des Lesers ausschlaggebend für die Bestückung sein soll. Diese Fragen verdeutlichen, dass das Newcomb-Paradox möglicherweise nur aufgrund der Verknüpfung unvereinbarer Annahmen, nicht aber wegen eines inneren Fehlers der orthodoxen Entscheidungstheorie zustande kommt. Theorien des rationalen Handelns sind ja für Situationen gemacht, in denen die Entscheider eine bedeutsame Wahl treffen können; nimmt man ihnen die freie Wahl, so entsteht eine Situation, über die

Selbstbezüglichkeit zu paradoxen Ergebnissen führen kann, ist seit dem Altertum bekannt³⁰ und treibt letztlich das Gödel-Theorem an.³¹

- Bei dem einmaligen Auftreten individueller Zeitinkonsistenz ist die ältere Inkarnation des Individuums – der Zeilenspieler – unwiderruflich untergegangen, wenn sein Nachfolger am Zug ist.
- Bei wiederholten Interaktionen, etwa den klassischen Problemen inkonsistenter Zeitpräferenzen oder bei Sucht, existieren die beiden Inkarnationen gleichsam parallel – wenn man so will, als intrapersonelle Regel- und Handlungsebene –, und es kommt auf die Rahmenbedingungen (oder auch das „Framing“, siehe Traub 1999) an, welche von ihnen aktiv ist.
- Bei mehreren Akteuren – also dem für die Politikanalyse relevanten Standardproblem der Zeitinkonsistenz, wie es nach Kydland und Prescott (1977) insbesondere in der Geldpolitik und der Theorie der Besteuerung untersucht wurde, befinden wir uns wieder im herkömmlichen Raum rein spieltheoretischer und strategischer Überlegungen

Elster behandelt diese Kategorien getrennt (allerdings auch weitgehend ohne Rückgriff auf formale Modellierung) und unterscheidet noch feiner innerhalb der oben genannten Fälle, etwa zwischen inkonsistenten Zeitpräferenzen und Suchtproblemen (Elster 1987). *Sein Schwerpunkt ist die Diskussion rein intrapersoneller Zeitinkonsistenz*, nicht die der zugrunde liegenden strategischen Situation allgemein.

In der Tat dürften materielle Unterschiede zwischen den verschiedenen intrapersonellen Konflikten bestehen, daher könnte es durchaus zulässig sein, den listenreichen Odysseus als vollständig rational einzustufen, obgleich man Strotz' (1955) Zentralfigur nur für beschränkt rational hält. Mit anderen Worten: Man könnte die Struktur der Präferenzen im Strotz-Fall für ein Merkmal eingeschränkter Rationalität halten, weil ein solches Individuum bei isolierten Entscheidungen und ohne Veränderung seiner Präferenzstruktur keinen konsistenten erstbesten Plan fassen kann, obwohl man an einer Konzeption vollständiger Rationalität festhält, bei der die Individuen zu jedem Zeitpunkt eine Entscheidung über Regelbefolgung und Regelbruch treffen können.

Der nächste Schritt baut auf der formalen Analogie zwischen dem Problem des Odysseus und einer interpersonalen Commitment-Spielsituation auf. McClennen (1993, S. 191) argumentiert nun, dass bei Auftreten eines solchen Zeitinkonsistenz-Problems genau wie im Fall des Odysseus das Individuum, welches die Selbstbindung installiert hat, sich nur selbst im Weg steht. Wiederum sei es nämlich der Mangel an Willenskraft des sich bindenden Spielers, der die Erfolglosigkeit des ersten Entschlusses bewirkt, und dabei handele es sich um ein rein internes Problem dieses Akteurs. Einem resoluten

Entscheidungstheorien weiter nichts aussagen können und die bei geeigneter Kombination von Annahmen jede solche Theorie überfordern.

³⁰ In Form des Epimenides-Paradoxes: Der Kreter Epimenides sagt, alle Kreter seien Lügner. Eine etwas komplexere Variante besteht in der Implikation: „Wenn dieser Satz wahr ist, dann bin ich Dracula.“ Die geneigte Leserin ist eingeladen, die Wahrheitstabelle für diesen Satz aufzustellen.

³¹ Wenn ein aussagelogisches System vollständig in dem Sinne ist, dass es alle denkbaren Aussagen formal abbilden kann, dann muss es auch die Formulierung selbstbezüglicher Aussagen gestatten. Mit diesen wiederum lassen sich ähnlich wie im Text und in der vorangegangenen Fußnote angesprochen Widersprüche konstruieren. Folglich kann ein vollständiges aussagelogisches System nicht widerspruchsfrei sein.

Individuum, das einmal gefasste Pläne bei unverändertem Datenkranz auch durchführt *und für diese seine Willenskraft bekannt ist*, gelingt es ohne Weiteres, die Selbstbindung zu verwirklichen, ohne dass es dazu auf externe Bindungsmechanismen zurückgreifen müsste. Die kursiv gesetzte Bedingung im letzten Satz ist wesentlich. Für das Verhalten des Zeilenspielers kommt es auf die Glaubwürdigkeit der Selbstbindung des Spaltenspielers an, und mangels eines externen Durchsetzungsmechanismus wird die Regel nur glaubwürdig sein, wenn der Zeilenspieler von der Willenskraft des Spaltenspielers weiß. Resoltheit allein genügt im interpersonell-strategischen Zusammenhang also nicht mehr, vielmehr muss noch ein allgemeines Wissen um die prozedurale Rationalität hinzutreten, das der oft kritisierten Annahme der *common knowledge of rationality* in der Spieltheorie (Hargreaves-Heap und Varoufakis 1995) analog ist.

Auf der formalen Seite ergeben sich noch tiefer greifende Änderungen. Wenn nämlich vergangene Planungen einen unmittelbaren Einfluss auf künftige Entscheidungen haben sollen, der über indirekte Wirkungen durch eine Veränderung der Nebenbedingungen oder durch das Aufrufen externer Bindungsmechanismen hinausgeht, kann das Entscheidungsproblem nicht mehr auf dem Weg der Rückwärtsinduktion gelöst werden. Denn dabei spielt die Existenz eines ex ante gefassten Plans an sich für die Entscheidungen in der Zukunft keine Rolle mehr (siehe erneut die Einleitung und die Diskussion im Abschnitt 1). Prozedurale Rationalität verlangt, dass in solchen Fällen die dynamische Programmierung als Entscheidungsverfahren als irrational aufgegeben wird, ausschlaggebend für das Handeln soll allein die ex ante vorzunehmende Gestaltung des optimalen Plans sein. Spieltheoretisch erfordert dies den Verzicht auf das Separabilitätsprinzip, nach dem man einzelne Teilspiele eines Spiels herausgreifen und unabhängig von den im Rest des Spiels getroffenen oder geplanten Entscheidungen analysieren kann, weil ein rationaler Spieler, der im Rahmen eines Spiels in dieses Teilspiel eintritt, sich ebenso verhält wie ein Spieler, der nur dieses Teilspiel *de novo* spielt.

Offenkundig bringt dies mit sich, dass man auch die Teilspielperfektheit (und *mutatis mutandis* das sequenzielle Gleichgewicht) als Lösungskonzept aufgeben muss, wenn man sich für diese Variante der prozeduralen Rationalität entscheidet. Aus zwei Gründen erscheint mir freilich eine gewisse Skepsis angebracht, was die Aufgabe des Separabilitätsprinzips anbelangt: Erstens halte ich bestimmte Aussagen, die sich bei der Zugrundelegung der Teilspielperfektheit ergeben – insbesondere über die Wirkungslosigkeit leerer Drohungen – für intuitiv plausibel. Allgemein dürfte es kaum Grenzen für die Drohungen geben, die ein resoluter Entscheider mit Aussicht auf Erfolg aussprechen könnte. Dies erscheint unbefriedigend. Zweitens vereinfacht die Anwendung der Rückwärtsinduktion das Kalkül des Entscheiders wesentlich. Ein resoluter Entscheider benötigt damit nicht nur erheblich mehr Willenskraft als ein „orthodox rationales“ Individuum, sondern c.p. auch eine größere kognitive Kapazität.

Unsere Systematik von Regelbegründungen und Regelbindungsproblemen (Abschnitt 2) hat bereits die Nähe von Halbdilemmata und gesellschaftlichen Dilemmata (PD-Situationen) gezeigt. Wenn man so will kann man sich diese letztlich als zwei ineinander verschränkte Halbdilemmata vorstellen (siehe dazu auch die Normalform in der Tab. 2). Daher könnte man in einem dritten Schritt verlangen, dass vollständige, wohlverstandene Rationalität auch Kooperation im nicht wiederholten Gefangenen-Dilemma verlange. Und in der Tat schreibt McClennen (1993, S. 195):

In such a game, however, a given player only faces an assurance problem in virtue of the conclusion that expected co-operation on his own part will trigger a decision to defect by the other player. That conclusion, in turn, pivots once again on the assumption that a rational player must be committed to the separability principle.

Doch anders als in Halbdilemmata liegen die Individuen hier selbst dann nicht ausschließlich im Streit mit sich selbst, wenn man dem zuvor skizzierten zweiten Schritt gefolgt ist.³² Eine Komponente des Dilemmas besteht sicherlich darin, dass jeder der Entscheider selbst versucht ist, den anderen durch Defektion auszubeuten. Da der andere dies weiß, ist er gut beraten, diese Ausbeutung zu vermeiden, indem er seinerseits defektiert. Ein kooperationswilliger Spieler hat somit das Problem, den anderen von seiner Kooperationsbereitschaft zu überzeugen, und dabei steht er sich bei geringer Willenskraft in der Tat nur selbst im Weg. Weil aber beide die Verlockung zum Regelbruch verspüren, hat jeder Spieler neben dem Problem mit sich selbst auch noch eines mit dem jeweils anderen: Es geht nicht nur darum, diesen von der eigenen Zuverlässigkeit zu überzeugen, sondern auch darum, nicht von ihm düpiert zu werden. Während ein prozedural rationaler Spieler Halbdilemma auch dann ohne externe Bindung die Vorteile der Regel realisiert, wenn er sich einem im üblichen Sinne rationalen Spieler gegenüber sieht, gelingt ihm das in einer PD-Situation nicht. Der Sinn der Regel im PD liegt nicht in der *Selbstbindung*, sondern in der Bindung *des anderen*.

Zweitens dürfte eine wesentliche Eigenschaft des Gleichgewichts in strikt dominanten Strategien verloren gehen, wenn man den dritten Schritt vollzieht, nämlich die Isolation des Spielers mit der dominanten Strategie vom Entscheidungskalkül seines Gegenübers. Die damit verbundenen informationellen und kognitiven Anforderungen sind offensichtlich: Das Individuum müsste letztlich die Rationalitätseigenschaften, Ziele und Restriktionen aller anderen kennen.³³ Bei reiner handlungsorientierter Rationalität spielen die Überlegungen, spielt die Rationalität des Gegenübers keine Rolle. Welche Handlung die anderen auch immer vornehmen (selbst wenn sie rein zufällig entscheiden), der betreffende Spieler fährt durch die Wahl der dominanten Strategie stets mindestens ebenso gut wie bei jeder anderen seiner Möglichkeiten.

Besonderes Gewicht dürfte diesen Einwänden drittens in Mehrpersonenspielen zukommen. Führt man die Logik dieser Argumentation konsequent weiter, könnte schließlich ein beliebiges Individuum in einer Welt mit vielen prozedural rationalen Entscheidern, selbst im Grenzfall des Atomismus, gesellschaftliche Variablen wie Preise nicht mehr als gegeben nehmen, sondern müsste einen optimalen Plan wählen,

³² Formal-spieltheoretisch könnte man sich diesem Argument wie folgt annähern. Jeder der beiden prozedural rationalen Akteure verfügt über mindestens drei Strategien: (1) Unbedingte Kooperation, (2) unbedingte Defektion, (3) kooperiere, wenn der andere Spieler ebenfalls Strategie (3) ankündigt, und defektiere sonst. (Daneben sind noch andere bedingte Strategien in Abhängigkeit von den Ankündigungen des jeweiligen Gegenübers denkbar.) In einer ersten Spielphase sollen nun die Spieler ihrem Gegenüber jeweils eine der drei Strategien versprechen. In der zweiten Spielphase werden die Strategien – u.U. in Abhängigkeit von der Ankündigung des Gegenübers – schlicht durchgeführt, wenn es sich um einen prozedural rationalen Spieler handelt; ein orthodox rationaler Spieler kann in der zweiten Spielphase zwischen den verfügbaren Handlungen (1) und (2) frei wählen. Bei prozeduraler Rationalität existiert die zweite Phase als eigenständiger Bestandteil des Spiels überhaupt nicht; bei orthodoxer Rationalität ist die erste Phase reiner „cheap talk“. Unter prozedural rationalen Spielern handelt es sich bei (3) um eine schwach dominante Strategie.

³³ Eine Konsequenz wäre, dass die Preise ihre Eigenschaft verlieren würden, besonders sparsame, aber zur Koordination ausreichende Signale zu sein. Ein vollständiges Ausschöpfen aller Tauschgewinne ohne externe Regelbindung und somit kollektives Handeln dürfte auch im Idealmodell nur gelingen, wenn die prozedural rationalen Entscheider unter ihresgleichen sind.

der die Reaktion aller anderen berücksichtigt – und auch beim Vorliegen öffentlicher Güter oder externer Effekte Tauschgewinne ausschöpft. Muss sich also Glazers und Konrads (1992) wohlhabendes Wirtschaftssubjekt, welches *ex ante* angesichts einer Einkommensunsicherheit für ein redistributives Steuer-Transfer-System votierte, bei vollständiger Rationalität *ex post* der Besteuerung unterwerfen und eine legale Steuervermeidung unterlassen? Aus einem sehr ähnlichen Grund plädiert Musgrave (1992, S. 376 f.) für eine wohlfahrtsökonomische Unbeachtlichkeit der *excess burden*, wenn man von einer sozialvertraglichen normativen Referenz ausgeht:

Having accepted Mill's rule of impartiality and joined in a contract to maximise aggregate welfare, individuals will also have subscribed to the more limited goal of minimising aggregate loss in distributing a given tax bill. Consistent therewith, they should stand ready to be taxed on what their incomes would be in the absence of tax, i.e. respond as if the tax was imposed in lump-sum form. There would then be no substitution effect and hence no deadweight loss. (...) That loss occurs in defiance of a contractual obligation to accept tax assignments that minimise aggregate loss, and thus loses standing. Rather than reward breach by crediting taxpayers with their dead-weight loss, that loss may be disregarded.

Musgrave geht hier allerdings noch nicht so weit zu sagen, dass rationale Individuen sich tatsächlich so verhalten müssten, als läge eine Pauschsteuer vor, sondern nur dass ihnen aus kontrakttheoretischer Sicht eine äquivalente Steuerzahlung abverlangt werden sollte. (Die Grenze der Ausbeutung der Reichen durch die Armen läge also im Aufkommensmaximum.) Sein Argument wird aber explizit in Kategorien von Regelbindungen und Regelbrüchen formuliert, und es ist mit der Vorstellung von resolutem Entscheiden nah verwandt.

Nachdem wir nun die einzelnen Schritte der Argumentation McClellens (1993) nachvollzogen und dabei bereits mögliche Einwände notiert haben, können wir die Diskussion nun mit dem Hinweis auf fundamentale Fragen schließen und die prozedurale Rationalitätskonzepte berühren. Das erste Problem hat mit dem Wesen einer Entscheidung zu tun – wenn man es volltönender mag, mit der Freiheit des Willens im Rahmen spieltheoretischer Modelle (Binmore 1994, S. 208 f.). Bislang habe ich nur bei jedem Schritt darauf hingewiesen, dass es titanischer Selbstbeherrschung und Willensstärke bedürfe, um die *ex ante* gefällte Entscheidung *ex post* gegenüber sich selbst durchzusetzen. Dass es aber überhaupt *möglich* ist, über *zukünftiges Handeln vorweg zu entscheiden*, wurde dabei stillschweigend unterstellt. Diese Prämisse wird indes häufig nicht ganz ohne Grund bestritten. So schreibt Vanberg (1994, S. 33):

We can make our choices only when they arise, we cannot make today our choices of tomorrow. Ulysses could not choose, before coming to the sirens, what he would do once he would be there. (...) In general terms, we can make choices only when they "come up".

Nach dieser Auffassung besteht insofern ein systematischer Unterschied zwischen Plänen und Absichten einerseits sowie Entscheidungen andererseits, als diese an die zeitlich bestimmte Tat geknüpft sind. Solange noch eine Möglichkeit zur Planrevision besteht, kann danach keine (endgültige) Entscheidung gefallen sein. Diese Sicht schließt die zeitliche Verlagerung von Entscheidungen zwar nicht aus, lässt sie aber nur in Form einer (Vor-)Verlagerung der Tat zu (Kliemt 1991). Dazu bedarf es der Installation einer wirksamen Regel, die dem Entscheider ein nachträgliches Abrücken vom Plan physisch unmöglich macht (oder stark verteuert), also dessen, was Elster (1979) und andere als „sophisticated choice“ bezeichnen. Aus dieser Perspektive formalisiert die Annahme sequenzieller Rationalität die *Verbindung zwischen Entscheidung und Tat* und ist daher notwendige Grundlage einer Theorie rationaler Entscheidungen; bei der Vorstellung,

ein Entscheider entschieße sich *ex ante*, *ex post* einer Regel zu folgen (siehe Rowe 1989, S. 22), handelt es sich dagegen um eine *contradictio in adiecto*.

4. Ratio und Emotionen: Warum ergreifen wir stets Partei für den Stackelberg-Führer?

“So I try hard to be good, \ And do the things I should, I give up everything that I enjoy. [...] All my friends say I’m a fool, \And I won’t break any rule \ While all the saintly virtues I employ [...]“

„But if the slightest fling \ Makes your soul begin to sing, \ That’s a necessary evil, my boy.“

Aus dem Jazz-Song “Necessary Evil”
(Evans)³⁴

„Wer die Kunst beherrscht, darf sie vernachlässigen.“

Thornton Wilder

Haben wir uns bislang mit den formalen Konsequenzen und der Möglichkeit befasst, vollständig rationale Entscheidungen so zu charakterisieren, dass das Auftreten von Zeitinkonsistenz ein Zeichen von Irrationalität und ihre Überwindung mittels externer Mechanismen ein Zeichen begrenzter Rationalität darstellt, so wollen wir uns nun der Frage zuwenden, ob es denn für den betroffenen Entscheider *wünschenswert wäre, eine vollständige, prozedural orientierte Rationalität zu erwerben*. Den Ausgangspunkt bildet also eine „Meta-Meta-Entscheidung“, in der ein rationales, vollständig informiertes Individuum seine eigene künftige *Disposition* wählt. Allerdings wird unsere Analyse insoweit rein exemplarischen Charakter tragen, als wir den zahlreichen kniffligen Fragen hinsichtlich dessen, was wir in einem derartigen Modell über die Persönlichkeit des Individuums – Präferenzen, Charakter und Disposition – als gegeben (in der Ausgangslage) betrachten können, im Folgenden aus Platzgründen ausweichen müssen. Bekanntlich kann man durch geeignete Annahmen über die *original position* jede konstitutionelle Lösung rationalisieren.

Mein Ziel in diesem Abschnitt ist denn auch bescheiden: Ich will verdeutlichen, dass es unter plausiblen Annahmen nicht notwendig zu einer Wahl der prozedural rationalen Disposition kommen muss. Damit führt prozedurale Rationalität womöglich zu einem Widerspruch, indem der perfekt rationale (im Sinne Elsters 1987 und McClennens 1993) Entscheider diese Disposition rational ablegte und lieber die indirekte Methode des Odysseus wählte. Den Schlüssel zu dieser Idee liefert wiederum

³⁴ Unsterblich in der Aufnahme mit Ella Fitzgerald und Louis Armstrong. Ausgelassen wurde der Text: “Well, a man sometime has to sip a bit (I’m hip I’m hip), cast his eyes on a lovely shade (Are you all right?), and then sometime, a wager he desires (If a man says he don’t, he’s not a man – he’s a liar. So don’t let yourself be sad, if what looks good is bad – reach out, accept the things that bring you joy).”

Elsters (1998; 2007) Werk, nun indes seine Arbeiten zum Verhältnis von Emotionen und Rationalität.

Zunächst wollen wir das Pferd von hinten aufzäumen und mit PD-Situationen beginnen. Hier ist man zwar gewöhnt, nach Möglichkeiten einer Überwindung zu suchen, doch hat es sich eingebürgert, auch Fälle zu betrachten, in denen man eine PD-Situation nach Möglichkeit kreieren oder ihre Überwindung erschweren sollte, weil diese zu Lasten Dritter ginge.³⁵ Bei den Halbdilemmata dagegen schlagen sich die meisten Autoren auf die Seite des jeweiligen Stackelberg-Führers, aus dessen Warte sie zu normativen Empfehlungen kommen. Im Fall strategischer Komplementarität mag dies wenig schaden,³⁶ doch kann man eine *allgemeine* Vernachlässigung der Position des Folgers normativ wohl kaum rechtfertigen. Was schließlich intrapersonale Konflikte angeht, ist mir *überhaupt kein* Autor bekannt, der explizit die Partei der folgenden Inkarnation ergriffe. Bewertet wird auf der Grundlage der vorherigen, langfristigen, überlegten, am konstitutionellen Interesse orientierten, auf der Meta-Ebene verorteten, normativ akzeptablen, „ethischen“, schlicht rationalen Präferenzen – oder welchen Duktus auch immer die jeweilige Autorin einzunehmen gedenkt. Der intrapersonale Stackelberg-Folger bleibt ein Störfaktor – versoffen, impulsiv, kurzichtig, kurzum: irrational.

In vielen Situationen, die durch Willensschwäche gekennzeichnet sind, kann diese Sichtweise, wie sie Elster intensiv diskutiert, durchaus angemessen sein. Das allein aber macht Willensstärke noch nicht zu einem Desiderat. Elster (1987, S. 72 f.) formuliert die zentrale Frage, allerdings leider nur mit Bezug auf die indirekte Selbstbindung:

Wenn jemand an einem Tag die Anweisung gibt, dass zukünftige Anweisungen einer bestimmten Art nicht zu befolgen seien, und dann am anderen Tag die Meta-Anweisungen zurücknimmt, worin bestünden dann die Entscheidungskriterien dafür, ob das Selbst, das die Rücknahme vorgenommen hat, authentischer ist als dasjenige, welches die Meta-Anweisungen gegeben hat, oder gar das Selbst, dessen Mangel an Authentizität diese Meta-Anweisungen notwendig gemacht hat?

Er liefert auch in seinen früheren Werken die wesentlichen Ansätze zu einer Skepsis an der einseitigen Parteinahme für den ursprünglich gefassten Plan:

Selbst wenn die Überlegungen, das Verhalten oder den Charakter zu ändern, den gewünschten Erfolg haben, kann die bloße Aktivität des Überlegens den Charakter zum Schlechteren verändern, in Bereichen, die als noch wichtiger betrachtet werden, indem sie nämlich die Spontaneität verkümmern lässt. Es ist möglich, von seiner Willenskraft abhängig zu werden, und es kann genau so rational sein, etwas gegen diese Abhängigkeit zu unternehmen, wie gegen jede andere. (Elster 1987, S. 72. Hervorhebung unterdrückt, KB.)

Für andere Menschen würde der Prozess, durch den sie auf den Gipfel gelangen (was sie wollen), auch zu einer Charakteränderung führen (wogegen sie sich sträuben); in diesem Fall könnte es rational sein, die Möglichkeit, deren Existenz man annimmt, zurückzuweisen. (Elster 1987, S. 77)

Die niedrigste Form moralischen Lebens ist vermutlich totale und kurzsichtige Impulsivität. Auf einer höheren Ebene steht ein Leben entsprechend den selbst auferlegten Regeln oder Strategien der Selbstverpflichtung; und auf der ausgesprochen höchsten Ebene der Bruch dieser

³⁵ Man denke an die Durchsetzung von Kartell-Absprachen und die Wettbewerbspolitik oder an die Mafia und die Strafverfolgung.

³⁶ Beim Gros der einschlägigen Arbeiten aus Finanztheorie und Theorie der Wirtschaftspolitik kommt hinzu, dass es sich bei dem Stackelberg-Führer um den „sozialen Planer“ handelt, dessen Zielfunktion (die soziale Wohlfahrtsfunktion) ohnedies den Maximanden in der normativen Theorie bildet.

Regeln, wenn dies unter Berücksichtigung aller Umstände gerechtfertigt erscheint. (Elster 1987, S. 136)

Hier ist alles Wesentliche für unsere Fragestellung angelegt, einschließlich der Trennung von Handlungs-, Meta- und Meta-Meta-Ebene sowie der ersten Gründe für eine Rechtfertigung eines Verzichts auf „vollständige“ Rationalität im zuvor (Abschnitt 3) beschriebenen Sinne – also der Frage, „dass ich manchmal gute Gründe für meine Bitte habe, entlassen zu werden, in dem Sinne, dass ich die Gründe akzeptiert hätte, bevor ich darum bat, gebunden zu werden“ (Elster 1987, S. 139). *Darin liegt meines Erachtens Elsters besonderer Beitrag zur Theorie der Zeitinkonsistenz.*³⁷

Leider wurde meines Wissens der hier aufkommende Gedanke, eine Skepsis gegenüber der vollkommenen Rationalität sturer Planbefolgung aus einem übergeordneten rationalen Entscheidungskalkül abzuleiten, kaum weiter betrachtet. Verfolgen wir ihn also noch ein wenig weiter! Der listenreiche Odysseus brachte das Kunststück fertig, den Sirenen zu entgehen und sie doch mit großem Genuss und Verlangen zu hören. Würde ihm dies gelingen, wenn er resolut wäre?

Als Alternative zu externen Bindungstechniken könnte Odysseus die Kosten der Planbefolgung grundsätzlich auf zweierlei Weise senken:

- Er könnte für eine gegebene Versuchung (Differenz des Periodennutzens bei Regelbruch und bei Regelbefolgung) die *psychischen Kosten des Regelbruchs erhöhen*, indem er in seiner Persönlichkeit ein Ethos der Willensstärke und Selbstüberwindung einführt, ausbaut und pflegt.
- Er könnte bei gegebenen psychischen Kosten, die Differenz des Periodennutzens verringern, indem er (a) seine Sensitivität für den Genuss verringert, der bei Regelbruch zu haben ist, und/oder (b) die Latenz künftiger Nutzen erhöht, sofern es sich um eine wiederkehrende Versuchung handelt.

Diese Strategieelemente stehen indes nicht unverbunden nebeneinander – eine stoische Persönlichkeit und einen buddhistischen Gleichmut erzeugt man traditionell *auch* dadurch, dass man Abstand von der Welt gewinnt und deren Reize abbaut, nicht durch das Sammeln von Willensstärke allein. Ebenso scheint die Fähigkeit, besonders intensive positive Emotionen zu empfinden, mit einer besonderen Intensität des Leidens einherzugehen. In seiner Auseinandersetzung mit Emotionen und deren Verhältnis zur Rationalität hat Elster wiederholt darauf hingewiesen, etwa in Elster (1989, S. 67),

If my life by and large is going well, a Stoic ability to endure adversity not only is pointless but is undesirable. To see why, note that one cannot choose emotions à la carte or, more generally, that emotional dispositions do not vary independently of each other. To ask for the ability to love without being vulnerable to grief is to ask for the moon, as is the desire to enjoy the euphoria of hope without being disappointed if the hoped-for event fails to occur. To be sure, “it’s all in the mind”, but the mind is not like a switchboard with one switch for each emotion. Because there are couplings between the emotions, a Stoic attitude comes at the cost of a flatter emotional life all around.

wobei seine Position im Zeitablauf vorsichtiger zu werden scheint:

Also, we do not understand whether and to what extent the propensity for emotional height goes together with the propensity for emotional lows. If it does, is a life in steady contentment

³⁷ Elster (op.cit.) spricht in diesem Zusammenhang auch das Problem eines infiniten Regresses von Meta-Ebenen kurz an. Sehr zu Recht! Siehe etwa Beckmann (2007a) oder Beckmann (2005).

more happy overall than one that alternates between euphoria and disphoria?³⁸ [...] The ideal of extinguishing the emotions that one finds in many ancient philosophies, notably Stoicism and Buddhism, emerged in societies where the environment may have offered more occasions for emotions with negative valence.

Wenn resolutes Entscheiden, wenn volle prozedurale Rationalität nicht ohne eine Verringerung der Genussfähigkeit – den stoischen „Gleichmut“ – zu haben sind, kann es freilich nicht mehr als ausgemacht gelten, dass unser „Meta-Meta-Entscheider“ sich für den stoischen Lebensplan entscheidet.³⁹ Möglicherweise fährt er als Odysseus, der seine große Empfänglichkeit für Reize und Genüsse mittels externer Mechanismen zähmt, im erwarteten Barwert besser. Wenn aber ein rationaler Entscheider, angesichts der verfügbaren Optionen bei seiner Lebens- und Selbstplanung, die Strategie des „*sophisticated choice*“ der resoluten Alternative bewusst vorziehen kann (was auch bedeutet, dass ein ursprünglich prozedural rationaler Entscheider diese Charaktereigenschaft womöglich rational *ablegen* wollte), kann man den Begriff der vollständigen Rationalität nicht mehr auf die letztgenannte Weise beschränken, mit dem Zeitinkonsistenz-Problem umzugehen. Der listenreiche Odysseus bleibt ein Kandidat für vollständige Rationalität.

5. Coda: Elster und das Zeitinkonsistenz-Problem

Selbstbindung und Zeitinkonsistenz sind zusammenhängende Themen von großer theoretischer und ordnungspolitischer Bedeutung. Unter anderem geht es um

- positive Fragen nach Mechanismen und Techniken zum Erreichen von Selbstbindung (das ist Schellings 1960; 1966; 2006 Hauptproblem, dem aber auch Elster 1979 breiten Raum gibt),
- normative verfassungsökonomische Fragen nach der Begründung von Regeln durch Vorteile einer Selbstbindung des Staates (die „rules vs. discretion“-Debatte, angestoßen von Kydland und Prescott 1977 sowie Barro und Gordon 1983),
- positive Fragen nach dem Charakter der Rationalität und der Falsifikation einer Theorie (voll-)rationaler Entscheidungen,
- normative Fragen nach unseren Anforderungen an einen *vollständig* rationalen Entscheider (und damit unserem Idealbild menschlicher Rationalität – ein Problem Elsters 1979 ebenso wie McClennens 1990),
- „meta-normative“ Fragen nach dem Stellenwert von tatsächlichem Handeln und tatsächlich empfundener Lebensqualität für die Würdigung von Wirtschaftspolitik. Denn wenn ein Individuum bei Rationalität in einer bestimmten Lebenslage etwas anderes wollen sollte, als es bei beschränkter

³⁸ Man beachte, dass die Vorstellung vom abnehmenden Grenznutzen hier nicht weiterhilft, denn der Grenznutzen des Nutzens ist konstant.

³⁹ Im engen Zusammenhang damit steht der Umgang mit „nicht nachhaltigen“ Genussmöglichkeiten. Ist es angesichts der Gewöhnungseffekte und der daher zu erwartenden Entzugsschmerzen rational, auf den Genuss zu verzichten, oder nicht?

Rationalität will, welchen Wert kann man dann seinen bekundeten Präferenzen noch beimessen?⁴⁰

Alle Fragen verbindet eine gemeinsame formale Grundstruktur (herausgearbeitet im Abschnitt 1) und der Umstand, dass Zeitinkonsistenz-Probleme nur die Rückseite einer Medaille darstellen, auf deren Vorderseite „Selbstbindung“ geprägt ist. Fraglich bleibt jedoch, inwieweit diese formale Analogie auch materiell trägt.⁴¹ Elster (2000) bleibt hier skeptisch. Und auch bei unserer eigenen Diskussion im Abschnitt 3 wurden Zweifel laut, ob man aufgrund formaler Analogien und Ähnlichkeiten tatsächlich mit McClennen (1990; 1993) einen Bogen von der Überwindung intrapersonaler Zeitinkonsistenz hin zur Kooperation im einfachen Gefangenen-Dilemma schlagen kann.

Andererseits sollte man das Kind nicht mit dem Bade ausschütten. Zum einen scheint mir die Vorstellung eines „Spiels gegen sich selbst“ beileibe nicht abwegig zu sein: Individuen, die ihre Autoschlüssel bei Barkeepern abgeben, die Abmachungen mit sich selbst treffen und sich für besondere Erfolge mit kleinen Genüssen belohnen, indem sie selbst auferlegte Regeln brechen, spielen solche Spiele. Zum anderen halte ich den Umstand, dass man zwei Phänomene mit dem gleichen formalen Modell (als eine Spielsituation) abbilden kann, für hochgradig signifikant, legt er doch eine strukturelle Verwandtschaft der zentralen Ursache-Wirkung-Beziehungen nahe. Elsters Neigung, lieber nach zahlreichen „Mechanismen“ als nach gemeinsamen Modellstrukturen zu suchen, scheint mir nicht der richtige Weg.⁴² Sie äußert sich auch bei Elsters früheren Analyse der Zeitinkonsistenz (Elster 1979; Elster 1987) in einem regelrechten *embarras de richesse* an Fällen.

Gleichwohl hat Jon Elster einen erheblichen Beitrag zur Theorie der Zeitinkonsistenz – und *a fortiori* der Selbstbindung – geleistet. Dieser enthält neben einer tiefen, zu Recht psychologisch geprägten Diskussion einzelner Phänomene auch ein Ringen um den Begriff menschlicher Rationalität, das durchaus aktuell geblieben ist, auch wenn Elster selbst nur mehr wenig Hoffnung auf den Rationalwahlansatz setzt. (Auch hierin mag ich ihm nicht folgen.) Der größte Teil von Elsters Beitrag zum Verständnis der Zeitinkonsistenz-Problematik besteht jedoch in etwas, was weder als solcher Beitrag intendiert noch eingeordnet wurde: in der Einführung einer dritten Ebene der *Entscheidungen über die eigene Persönlichkeit* und dem Gedanken, dass es auf dieser Ebene nicht rational sein muss, im Elster'schen (1979; 1987) Sinne vollständig rational zu sein.

⁴⁰ Es hilft freilich wenig, dass alle Spekulationen über das „wirklich Gewollte“ kontrafaktisch sein müssen und ihre Zulässigkeit im wirtschaftspolitischen Diskurs zudem das Tor weit für Manipulationen und Paternalismen öffnet. Für mein Plädoyer, bei einer konstitutionell orientierten Wirtschaftspolitik nur die empirisch eruierten, wirklichen konstitutionellen Präferenzen realer Individuen einfließen zu lassen, siehe Beckmann (1998, Kap. 3), für einen Ansatz, wie man dies praktisch bewerkstelligen könnte, siehe die Bemerkungen bei Brennan und Buchanan (1993).

⁴¹ Dieser Frage widmeten sich große Teile der Debatte während des Symposiums.

⁴² Diese Vorgehensweise bleibt vermutlich dem Streben nach Erklärung des Einzelfalls geschuldet, das Elsters Werk kennzeichnet.

Literatur

- Robert J. Barro/D. B. Gordon, Rules, discretion and reputation in a model of monetary policy (Journal of Monetary Economics) (1983).
- Klaus Beckmann, Analytische Grundlagen einer Finanzverfassung (Frankfurt am Main 1998).
- Klaus Beckmann, Konstitutionenökonomik – Ein Theorieangebot für die Politikberatung? In: Martin Leschke, Ingo Pies (Hrsg.), Wissenschaftliche Politikberatung – Theorien, Konzeptionen, Institutionen (Stuttgart 2005), S. 376–399.
- Klaus Beckmann, Thomas Schelling und die Militärstrategie. In: Martin Leschke, Ingo Pies, Thomas Schellings strategische Ökonomik (Tübingen 2007a), S. 189–216.
- Klaus Beckmann, Europa – Feind der Freiheit? Betrachtungen zu Harmonisierung und Wettbewerb. In: Wolf Schäfer, Andreas Graf Wass von Czege (Hrsg.), Das Gemeinsame Europa – viele Wege, kein Ziel? (Baden-Baden 2007b).
- Ken Binmore, Game Theory and the Social Contract, Bd. 1 (Cambridge, Mass. 1994).
- Ken Binmore, Playing for Real (Oxford 2006).
- Geoffrey Brennan/James M. Buchanan, Die Begründung von Regeln (Tübingen 1993).
- Antony de Jasay, Social Contract, Free Ride: A Study of the Public Goods Problem (Oxford 1989).
- Jon Elster, Ulysses and the Sirens (Cambridge, UK 1979).
- Jon Elster, Sour Grapes. Studies in the Subversion of Rationality (Cambridge, UK 1985).
- Jon Elster (Hrsg.), Rational Choice (New York 1986).
- Jon Elster, Subversion der Rationalität (Frankfurt/New York 1987).
- Jon Elster, Alchemies of the Mind. Rationality and the Emotions (Cambridge, UK 1998).
- Jon Elster, Ulysses Unbound. Studies in Rationality and Irrationality (Cambridge, UK 2000).
- Jon Elster, Explaining Social Behavior. More Nuts and Bolts for the Social Sciences (Cambridge, Mass. 2007).
- Esther Gal-Or, First mover and second mover advantages (International Economic Review, 26) (1985), S. 649–653.
- David Gauthier, Maximization constrained: the rationality of cooperation. In: Richmond Campbell/Lanning Snowden (Hrsg.), Paradoxes of Rationality and Cooperation (Vancouver 1985).
- Amihai Glazer, Politics and the choice of durability (American Economic Review, 79) (1989).
- Amihai Glazer/Kai Konrad, Intertemporal Commitment Problems and Voting on Redistributive Taxation (Arbeitspapier Sonderforschungsbereich, 178) (1992).
- Shaun Hargreaves-Heap/Yanis Varoufakis, Game Theory. A Critical Introduction (London/New York 1995).
- Mark Harrison, An economist looks at suicide terrorism (World Economics, 7) (2006), S. 1–15.
- John C. Harsányi, Rule utilitarianism, equality, and justice. In: Ellen Frankel Paul, Fred D. Miller, Jeffrey Paul, Ethics and Economics (Oxford 1985).
- Friedrich A. von Hayek, Die Verfassung der Freiheit (Tübingen 1991).
- Ronald A. Heiner, The origin of predictable behavior (American Economic Review, 73) (1983), S. 560–595.
- Ronald A. Heiner, Rule-governed behavior in evolution and human society (Constitutional Political Economy, 1) (1990), S. 19–42.

- Christiansen Jensen, Optimal implementation delays: when should policies be announced? (*Scottish Journal of Political Economy*, 54/4) (2007), S. 492–507.
- Hartmut Kliemt, Der Homo Oeconomicus in der Klemme. In: H. Esser, G. Troitzsch (Hrsg.), *Modellierung sozialer Prozesse* (Bonn 1991).
- Finn E. Kydlandund/Edward C. Prescott, Rules rather than discretion: the inconsistency of optimal plans (*Journal of Political Economy*, 85) (1977), S. 473–491.
- Bernd Lahno, *Versprechen. Überlegungen zu einer künstlichen Tugend* (München/Wien 1995).
- Bernd Lahno, *Der Begriff des Vertrauens* (Paderborn 2002).
- David M. Levy, *The Economic Ideas of Ordinary People: From Preferences to Trade* (New York 1992).
- Aldo Matteuci, Is suicide terrorism a novel economic phenomenon? (*World Economics*, 8) (2007), S. 239–240.
- Edward F. McClennen, *Rationality and Dynamic Choice* (Cambridge, UK 1990).
- Edward F. McClennen, Rationality, constitutions, and the ethics of rules (*Constitutional Political Economy*, 4) (1993), S. 173–207.
- Richard A. Musgrave, Social contract, taxation and the standing of deadweight loss (*Journal of Public Economics*, 49) (1992), S. 369–381.
- Ingo Pies, *Ordnungspolitik in der Demokratie: ein ökonomischer Ansatz diskursiver Politikberatung* (Tübingen 2000).
- Ingo Pies, *Wissenschaftliche Politikberatung in der Demokratie*. In: Klaus Beckmann, Dietmar Meyer, Stefan Okruch (Hrsg.), *Neuer Wein aus alten Schläuchen?* (Budapest 2003).
- Nicolas Rowe, *Rules and Institutions* (Ann Arbor 1989).
- Thomas J. Sargent, *Dynamic Macroeconomic Theory* (Cambridge, Mass. 1987).
- Thomas C. Schelling, *The Strategy of Conflict* (Cambridge, Mass. 1960).
- Thomas C. Schelling, *Arms and Influence* (New Haven 1966).
- Joseph A. Schumpeter, *Kapitalismus, Sozialismus und Demokratie* (Tübingen 2005, 8. Aufl., Originalausgabe 1942).
- Robert H. Strotz, Myopia and inconsistency in dynamic utility maximization (*Review of Economic Studies*, 23/3) (1955), S. 165–180.
- Stefan Traub, *Framing Effects in Taxation* (Heidelberg 1999).
- Viktor J. Vanberg, *Rules and Choice in Economics* (London/New York 1994).

Diskussionspapiere⁴³

- Nr. 2007-21 **Klaus Beckmann**
Jon Elster und das Zeitinkonsistenz-Problem
- Nr. 2007-20 **Ingo Pies, Stefan Hielscher**
Nachhaltigkeit in Forschung und Lehre – Die „Principles for Responsible Management Education“ des UN Global Compact
- Nr. 2007-19 **Claus Offe**
Die genetische Entschlüsselung der politischen Ordnung □ Jon Elster über Selbstbindung und die Verfassungsprozesse in Osteuropa
- Nr. 2007-18 **Ingo Pies, Jörg Viebranz**
Jon Elster und das Problem lokaler Gerechtigkeit
- Nr. 2007-17 **Ingo Pies**
Theoretische Grundlagen demokratischer Wirtschafts- und Gesellschaftspolitik – Der Beitrag von Jon Elster.
- Nr. 2007-16 **Ingo Pies, Markus Beckmann, Stefan Hielscher**
Mind the Gap! – Ordonomische Überlegungen zur Sozialstruktur und Semantik moderner Governance
- Nr. 2007-15 **Ingo Pies, Stefan Hielscher, Markus Beckmann**
Theorie für die Praxis – Zehn Thesen zur wirtschafts- und unternehmensethischen Ausbildung von Führungskräften
- Nr. 2007-14 **Sabine Campe, Lothar Rieth**
Wie können Corporate Citizens im Global Compact voneinander lernen? Bedingungen, Hemmnisse und Bewertungskriterien
- Nr. 2007-13 **Ingo Pies, Stefan Hielscher, Markus Beckmann**
Wie können Corporate Citizens voneinander lernen? Anregungen für den Global Compact der Vereinten Nationen
- Nr. 2007-12 **Ingo Pies**
Unternehmensethik in der Marktwirtschaft: Moral als Produktionsfaktor
- Nr. 2007-11 **Ingo Pies**
Fragen und Antworten – Wie reagiert eine ökonomische Theorie der Moral auf die „Frequently Asked Questions“ zur Wirtschafts- und Unternehmensethik?
- Nr. 2007-10 **Johanna Brinkmann-Braun, Ingo Pies**
The Global Compact’s Contribution to Global Governance Revisited
- Nr. 2007-9 **Ingo Pies, Stefan Hielscher**
The International Provision of Pharmaceuticals: A Comparison of Two Alternative Theoretical Strategies for Global Ethics
- Nr. 2007-8 **Markus Beckmann**
NePAD und der African Peer Review Mechanism – Zum Potential politischer Selbstbindung
- Nr. 2007-7 **Karl Homann**
Moral oder ökonomisches Gesetz?
- Nr. 2007-6 **Markus Beckmann, Ingo Pies**
Responsibility and Economics
- Nr. 2007-5 **Ingo Pies, Stefan Hielscher**
Das Problem weltmarktlicher Arzneimittelversorgung:
Ein Vergleich alternativer Argumentationsstrategien für eine globale Ethik
- Nr. 2007-4 **Ingo Pies**
Wie bekämpft man Korruption? – Lektionen der Wirtschafts- und Unternehmensethik

⁴³ Als Download unter <http://ethik.wiwi.uni-halle.de/forschung>. Hier finden sich auch die Diskussionspapiere der Jahrgänge 2003 und 2004.

- Nr. 2007-3 **Markus Beckmann**
Ökonomischer Ansatz und die Theorie des Self-Command bei Thomas Schelling
- Nr. 2007-2 **Ingo Pies**
Theoretische Grundlagen demokratischer Wirtschafts- und Gesellschaftspolitik – Der Beitrag von Thomas Schelling
- Nr. 2007-1 **Ingo Pies, Stefan Hielscher**
Das Problem der internationalen Arzneimittelversorgung: Eine wirtschaftsethische Perspektive
- Nr. 2006-13 **Ingo Pies**
Ist Konsens im Konflikt möglich? – Zur gesellschaftstheoretischen und gesellschaftspolitischen Bedeutung von Metaspielen
- Nr. 2006-12 **Ingo Pies**
Nachhaltigkeit: eine semantische Innovation von welthistorischer Bedeutung
- Nr. 2006-11 **Markus Beckmann, Ingo Pies**
Ordo-Responsibility – Conceptual Reflections towards a Semantic Innovation
- Nr. 2006-10 **Markus Beckmann, Ingo Pies**
Ordnungsverantwortung – Konzeptionelle Überlegungen zugunsten einer semantischen Innovation
- Nr. 2006-9 **Markus Beckmann, Ingo Pies**
Freiheit durch Bindung – Zur ökonomischen Logik von Verhaltenskodizes
- Nr. 06-8 **Ingo Pies**
Methodologischer Hobbesianismus und das Theorieprogramm einer interessenbasierten Moralbegründung
- Nr. 06-7 **Ingo Pies, Peter Sass**
Korruptionsprävention als Ordnungsproblem – Wirtschaftsethische Perspektiven für Corporate Citizenship als Integritätsmanagement
- Nr. 06-6 **Nikolaus Knoepffler**
Projekt „Wirtschaftsethik“. Zur Aufhebung des Ansatzes von Karl Homann
- Nr. 06-5 **Ingo Pies, Alexandra von Winning**
Die UN-Dekade „Bildung für eine nachhaltige Entwicklung“ und der Beitrag des Global Compact: Eine wirtschaftsethische Perspektive
- Nr. 06-4 **Ingo Pies**
Markt versus Staat? – Über Denk- und Handlungsblockaden in Zeiten der Globalisierung
- Nr. 06-3 **Markus Beckmann, Diana Kraft, Ingo Pies**
Freiheit durch Bindung - Zur Logik von Verhaltenskodizes
- Nr. 06-2 **Sören Buttkereit, Ingo Pies**
The Economics of Social Dilemmas
- Nr. 06-1 **Ingo Pies**
Theoretische Grundlagen demokratischer Wirtschafts- und Gesellschaftspolitik – Der Beitrag von Albert Hirschman
- Nr. 05-12 **Ingo Pies, Alexandra von Winning**
Sustainability by Education: Lessons to be learned
- Nr. 05-11 **Ingo Pies, Alexandra von Winning**
Nachhaltigkeit durch Bildung: Lessons to be learned
- Nr. 05-10 **Ingo Pies**
Ökonomische Ethik: Zur Überwindung politischer Denk- und Handlungsblockaden
- Nr. 05-9 **Ingo Pies**
Chancen und Risiken der Globalisierung: 10 Thesen
- Nr. 05-8 **Gerhard Engel**
Karl Marx und die Ethik des Kapitalismus

- Nr. 05-7 **Ingo Pies**
Was gefährdet die Demokratie? – Eine kritische Stellungnahme zur Kapitalismusdebatte in Deutschland
- Nr. 05-6 **Martin Petrick, Ingo Pies**
In Search for Rules that Secure Gains from Cooperation: The Heuristic Value of Social Dilemmas for Normative Institutional Economics
- Nr. 05-5 **Stefan Hielscher, Ingo Pies**
Internationale Öffentliche Güter – Ein neues Paradigma der Entwicklungspolitik?
- Nr. 05-4 **Ingo Pies, Peter Sass**
Selbstverpflichtung als Instrument der Korruptionsprävention bei Infrastrukturprojekten
- Nr. 05-3 **Ingo Pies**
Theoretische Grundlagen demokratischer Wirtschafts- und Gesellschaftspolitik – Der Beitrag von Karl Marx
- Nr. 05-2 **Ingo Pies, Markus Sardison**
Wirtschaftsethik
- Nr. 05-1 **Johanna Brinkmann, Ingo Pies**
Corporate Citizenship: Raison d'être korporativer Akteure aus Sicht der ökonomischen Ethik

Wirtschaftsethik-Studien⁴⁴

- Nr. 2007-1 **Markus Beckmann**
Corporate Social Responsibility und Corporate Citizenship
- Nr. 2005-3 **Ingo Pies, Peter Sass, Roland Frank**
Anforderungen an eine Politik der Nachhaltigkeit – eine wirtschaftsethische Studie zur europäischen Abfallpolitik
- Nr. 2005-2 **Ingo Pies, Peter Sass, Henry Meyer zu Schwabedissen**
Prävention von Wirtschaftskriminalität: Zur Theorie und Praxis der Korruptionsbekämpfung
- Nr. 2005-1 **Valerie Schuster**
Corporate Citizenship und die UN Millennium Development Goals:
Ein unternehmerischer Lernprozess am Beispiel Brasiliens
- Nr. 2004-1 **Johanna Brinkmann**
Corporate Citizenship und Public-Private Partnerships: Zum Potential der Kooperation zwischen Privatwirtschaft, Entwicklungszusammenarbeit und Zivilgesellschaft

⁴⁴ Als Download unter <http://ethik.wiwi.uni-halle.de/forschung>.