

# Rationalität und Moralität

## Zum systematischen Ort der Moral im Kontext von Präferenzen und Restriktionen\*

GERHARD MINNAMEIER\*\*

Im Zentrum des Beitrags steht die Frage, wie individuelle Moralität im Kontext der Theorie rationalen Handelns und der Verhaltensökonomik konzipiert und modelliert werden kann. Zentral ist dabei die systematische Differenzierung zwischen Präferenzen und Restriktionen (im Sinne von Beliefs). Im ersten Teil wird gezeigt, dass dies im Rahmen bekundeter Präferenzen (revealed preferences) nicht bzw. nur sehr eingeschränkt möglich ist. Im zweiten Teil wird auf Basis eines aktuellen Ansatzes, der „Reason-based Theory of Rational Choice“, und einer neo-Kohlberg'schen Moraltheorie herausgearbeitet, wie dem genannten Anspruch Rechnung getragen werden kann.

Schlagwörter: Rational-Choice-Theorie, soziale Präferenzen, Homo oeconomicus, Framing-Effekte, moralische Motivation, situationsspezifische Moral, Neo-Kohlberg-Theorie

### ***Rationality and Morality – On the Systematic Place of Morality in the Context of Preferences and Beliefs***

*The contribution focuses on the question how individual morality can be conceptualised and modelled in the context of rational choice theory and behavioural economics. In this respect it is crucial to differentiate systematically between preferences, on the one hand, and restrictions (in terms of beliefs), on the other. The first part reveals that this is not – or only to some very limited extent – possible on the basis of so-called revealed preferences. In the second part, individual morality is reconstructed and analysed in the framework of a current approach called “reason-based theory of rational choice” and a neo-Kohlbergian moral theory.*

*Keywords: Rational Choice Theory, Reason-based Theory, Social Preferences, Homo Economicus, Framing Effects, Moral Motivation, Situation-specific Morality, Neo-Kohlbergian Moral Theory*

## 1. Problemstellung

In der verhaltensökonomisch orientierten Wirtschaftsethik stellt sich die grundlegende Frage, wie man moralisches Handeln erklärt bzw. modelliert. Diese auf den ersten Blick

---

\* Beitrag eingereicht am 23.01.2015; nach doppelt verdecktem Gutachterverfahren überarbeitete Fassung angenommen am 28.09.2015.

\*\* Prof. Dr. Gerhard Minnameier, Goethe Universität Frankfurt am Main, FB2: Wirtschaftswissenschaften, Professur für Wirtschaftsethik und Wirtschaftspädagogik, Theodor-W.-Adorno-Platz 4, D-60629 Frankfurt am Main, Tel.: +49-(0)69-798-34689, Fax: +49-(0)69798-35008, E-Mail: minnameier@econ.uni-frankfurt.de, Forschungsschwerpunkte: Moralische Urteilen und Handeln, Inferentielles Denken, Entwicklung individueller Moral und philosophischer Ethik, Ethik und Ökonomik.

einfache Frage erweist sich bei näherem Hinsehen als sperrig, und zwar aus zwei Gründen: *Erstens* ist zu klären, was (bestimmte Arten von) Moral kennzeichnet und wie individuelle Moralvorstellungen in Handeln umgesetzt werden können. Die so verstandene Fragestellung ist durchaus komplex. *Zweitens* ist sie zudem aus ökonomischer Perspektive schwer zugänglich. Das liegt nicht etwa daran, dass Ökonomik mit Moral oder Ethik nichts zu tun hätte. Vielmehr besteht die Schwierigkeit darin, der Moral im Rahmen (verhaltens-)ökonomischer Modellierungen einen systematischen Ort zuzuweisen. Diese Problematik ist in dem seit der paretianischen Wende eingeschlagenen Weg der ökonomischen Modellierung rationalen Handelns begründet (vgl. hierzu insbes. Hollis/Sugden 1993; Bruni/Sugden 2007).

Nach der Rational-Choice-Theorie (RCT) werden nutzenmaximierende Handlungen bzw. Entscheidungen über Präferenzen und Restriktionen modelliert. Jedes Individuum hat demnach eine subjektive Präferenzordnung und handelt jeweils unter bestimmten Restriktionen (klassischerweise Budgetrestriktionen). Obwohl als subjektiv konzipiert, werden Präferenzen seit Pareto aber nicht mehr auf psychologische Entitäten zurückgeführt und entsprechend erklärt, sondern als Rangordnungen potentieller Handlungsergebnisse (z.B. in Form von Güterbündeln) modelliert. Diese Präferenzen werden ihrerseits aus den faktischen Entscheidungen der Individuen (und unter spezifischen Rationalitätsannahmen) abgeleitet, weshalb wir seit Samuelson (vgl. 1938; 1948) von sogenannten „revealed preferences“ sprechen.

Problematisch wird diese Konstruktion jedoch, wenn man Handlungsrestriktionen nicht mehr auf objektive (Budget-)Restriktionen beschränkt, sondern alle relevanten, *subjektiv wahrgenommenen* Handlungsrestriktionen mit einbezieht, auf die Akteure bei ihren Verhaltensanpassungen reagieren (vgl. Bruni/Sugden 2007; Vanberg 2008). Seit Leonard Savage (vgl. 1954) werden auch Restriktionen als subjektiv aufgefasst und entsprechend als „beliefs“ modelliert (bei ihm insbesondere im Sinne subjektiver Wahrscheinlichkeiten). Die im Grunde einfache Vorstellung, dass Akteure ihren Nutzen gemäß ihren (stabilen) Präferenzen und den wahrgenommenen (situationsspezifischen) Restriktionen maximieren, führt allerdings zu dem Grundproblem, dass man von Entscheidungen nicht mehr im strengen Sinne auf Präferenzen zurückschließen kann, wie das nach dem Modell der „revealed preferences“ geschieht. Präferenzen und Restriktionen müssten nämlich voneinander unabhängig erfasst werden (können).

In der mikroökonomischen Modellierung behilft man sich, indem man für Präferenzen Rationalitätsannahmen macht, Restriktionen (experimentell) möglichst streng kontrolliert bzw. konstant hält und so gleichwohl Präferenzen zu extrahieren versucht. Allerdings sind auch die Rationalitätsannahmen keineswegs analytisch und insofern alles andere als unproblematisch (vgl. Bruni/Sugden 2007: 169–171). Gleiches gilt für Annahme der strikten Stabilität von Präferenzen, denn wären sie erfüllt, dann könnten sich grundlegende Präferenzen niemals – z.B. infolge moralischen Lernens – verändern oder entwickeln (ebd.: 164f.).

Noch schwerwiegender wird das skizzierte Problem jedoch, wenn man bedenkt, dass sich die Moral nicht in der Bevorzugung bestimmter konkreter Handlungsergebnisse ausdrückt und schon gar nicht erschöpft. Wenn man z.B. im Diktatorspiel dem passiven Spielpartner einen positiven Betrag zuweist, dann hängt die Frage der Moral nicht vom Ergebnis ab, sondern vom Handlungsmotiv, sprich dem moralischen Grund, aus dem

heraus man etwas abgibt und wie viel man gibt.<sup>1</sup> Für solche Gründe aber ist in der standardökonomischen Modellierung bislang kein Platz. Dietrich und List (2013a; 2013b) haben allerdings jüngst eine „Reason-based Theory of Rational Choice“ (RBT) vorgeschlagen, auf deren Basis man dieses Problem eventuell lösen könnte.

Vor diesem Hintergrund soll im vorliegenden Beitrag untersucht werden, wie individuelle Moralität im Rahmen der Theorie rationalen Handelns systematisch verortet und modelliert werden kann. Im nachfolgenden *Abschnitt 2* wird zunächst über mehrere Stationen herausgearbeitet, inwieweit Moralität überhaupt in einer Rational-Choice-Modellierung einen systematischen Ort finden kann (2.1), ob Moralität im Bereich von Präferenzen oder Restriktionen anzusetzen wäre (2.2), wie mit Bezug auf Präferenzen und Restriktionen moralrelevante Framing-Effekte zu erklären sind (2.3) und wie demzufolge sogenannte „soziale Präferenzen“ und ein weiter ökonomischer Vorteilsbegriff im RCT-Rahmen systematisch verortet werden könnten (2.4).

Wie sich zeigt, scheint hierfür vor allem die bereits erwähnte RBT von Dietrich und List geeignet zu sein, die als Basis für den konstruktiven Ansatz in *Abschnitt 3* dient und zunächst in ihren Grundzügen vorgestellt wird (3.1). In *Abschnitt 3.2* wird dargelegt, wie die in der RBT zentralen „Gründe“ für den Bereich der Moral zu fassen wären. Dabei wird auf eine neo-Kohlberg'sche Moralstufentaxonomie zurückgegriffen. In *Abschnitt 3.3* wird erörtert, wie aus moralischen Prinzipien situationsspezifisch konkrete moralische Intentionen generiert werden. Und in *Abschnitt 3.4* geht es um die Frage der Umsetzung von Intentionen in Handeln und in diesem Zusammenhang auch um die Frage, inwiefern irrationales Handeln möglich ist bzw. im Rahmen der RBT modelliert werden kann.

## 2. Präferenzen und Restriktionen in der Rational-Choice-Theorie (RCT)

### 2.1 Der Homo Oeconomicus und die Moral

In der ökonomischen Analyse ist man längst vom Konzept des Homo Oeconomicus (HO) als eines Menschenbildes abgekommen. Spätestens in den 1970er Jahren begann die moderne Ökonomik, sich von dieser engen Modellierung menschlichen Handelns zu verabschieden (vgl. Arrow 1972; Becker 1976; Sen 1977; Akerlof 1982). Die neue, umfassende Perspektive hat Becker deutlich zum Ausdruck gebracht:

„The economic approach I refer to does not assume that individuals are motivated solely by selfishness or material gain. It is a method of analysis, not an assumption about particular motivations. Along with others, I have tried to pry economists away from narrow assumptions about self-interest. Behavior is driven by a much richer set of values and preferences“ (1993: 385).

Was Becker mit seinem Ansatz zu erreichen versucht, ist eine umfassende ökonomische Erklärung jedweden menschlichen Verhaltens, indem er faktische Verhaltensregularitäten auf subjektive Präferenzen und subjektiv wahrgenommene Restriktionen zurückführt. Üblicherweise spricht man in diesem Zusammenhang vom Rationalitätsprinzip –

---

<sup>1</sup> Das gilt selbst für utilitaristische Nutzenkalküle, die sich ja im Normalfall nicht nur auf die eine unmittelbare Handlungskompetenz beschränken.

im Unterschied zur Rationalitätsannahme – (vgl. Popper 1967/1997; Vanberg 2008) und versteht damit Rationalität bzw. den HO als Analyseschema, mit dessen Hilfe Verhalten in seiner *impliziten Rationalität* rekonstruiert wird (vgl. Homann/Suchanek 2005; Pies/Hielscher 2014).

Kirchgässner formuliert entsprechend das – wie er es nennt – „schwache Rationalitätsprinzip“: „Für jede beobachtbare Handlung ist eine Erklärung möglich, die diese als rationales Verhalten erklärt“ (2013: 262). Abgesehen davon, dass gerade die Intentionalität von Handlungen nicht beobachtbar ist (und daraus resultieren ja gerade die Rekonstruktionsprobleme),<sup>2</sup> verbinden sich mit diesem Rationalitätsprinzip jedoch mindestens zwei Probleme (vgl. hierzu auch Nickamp 2014): *Erstens* ist Irrationalität damit per se ausgeschlossen. Die Unterscheidung von Rationalität und Irrationalität wäre aber sowohl unter dem Aspekt menschlicher Fehlbarkeit als auch unter dem Aspekt normativer Wirtschaftsethik von besonderer Bedeutung.<sup>3</sup> *Zweitens* verläuft die RCT-basierte Erklärung über Präferenzen und Restriktionen, was bedeutet, dass prima facie „irrationales Verhalten“ *entweder* unter Rekurs auf spezifische Präferenzen *oder aber* unter Rekurs auf spezifische Restriktionen „rationalisiert“ werden kann. Damit bleibt freilich die Frage offen, an welcher „Stellschraube“ die Erklärung ansetzen soll.

## 2.2 Präferenzen oder Restriktionen – das ist hier die Frage

In einer Vielzahl ökonomischer Experimente konnte belegt werden, dass Menschen nicht im strengen Sinne eigeninteressiert handeln, sondern sich an Prinzipien der Fairness orientieren (vgl. insbes. Henrich et al. 2004; Fehr/Schmidt 2006). Dies führte zu verschiedenen Konzepten „sozialer Präferenzen“, mit deren Hilfe das Verhalten modelliert wurde (vgl. hierzu ebenfalls den Überblicksartikel von Fehr/Schmidt 2006). Parallel dazu wurde allerdings auch deutlich, dass die moralischen Verhaltenstendenzen anfällig für situationale Einflüsse waren. So fanden Hoffman et al. (vgl. 1996), dass die Fairnessorientierung abnimmt, wenn im Diktatorspiel eine Sprache verwendet wird, die das Teilen weniger insinuiert als die Instruktion im Standardexperiment. Die Bereitschaft etwas abzugeben, nimmt außerdem ab, wenn Anonymität und soziale Distanz zu den Rezipienten zunehmen. Im Extremfall sinkt die Bereitschaft zu teilen auf Werte nahe Null.

Diese Fragilität moralischer Handlungstendenzen wurde auch in weiteren Experimenten offensichtlich. Simple Veränderungen der Entscheidungsmöglichkeiten können erstaunliche Effekte zeitigen. List (vgl. 2007) replizierte zunächst in einem „Baseline-Setting“ des Diktatorspiels die Ergebnisse von Forsythe et al. (vgl. 1994), wonach etwa 20 Prozent der Teilnehmer die Hälfte ihrer Anfangsausstattung abgeben und nur ca. 20 Prozent alles für sich behalten (der Rest gibt einen dazwischen liegenden Betrag ab). Bei List bekamen beide Spieler zunächst \$5 und der Diktator noch einmal zusätzlich \$5, sodass hälftiges Teilen bedeutet, dass der Diktator \$2,50 transferiert. Räumt man den Diktatoren allerdings die Möglichkeit ein, dem Rezipienten auch etwas wegzunehmen,

<sup>2</sup> Beobachtbar und beobachtungssprachlich beschreibbar ist nur Verhalten. Handlungen implizieren hingegen Intentionalität, und diese muss bei der Analyse des Handelns erschlossen werden.

<sup>3</sup> Hierauf komme ich in Abschnitt 3.4 zurück.

dann erhält man ganz andere Ergebnisse, obwohl die Situation – so möchte man jedenfalls meinen – die gleiche ist. Besonders in der symmetrischen Variante, in der die Diktatoren den Rezipienten entweder bis zu \$5 geben oder umgekehrt bis zu \$5 nehmen konnten, reduzierte sich die Bereitschaft zu teilen drastisch. Nur noch 10 Prozent der Diktatoren gaben überhaupt etwas ab, dann allerdings in aller Regel die Hälfte (im Durchschnitt 42 Prozent der zusätzlichen \$5). Da die Struktur des Spiels im Wesentlichen die gleiche ist – es kommen ja nur scheinbar irrelevante Handlungsmöglichkeiten hinzu –, handelt es sich offenbar um einen Framing-Effekt, durch den die Norm des fairen Teilens weniger salient wird.

Aber nicht nur solche Framing-Effekte beeinflussen die Entscheidung, sondern auch die Möglichkeit, moralisch problematische Entscheidungen an Dritte abgeben oder kaschieren zu können (wie ja auch schon von Hoffman et al. (vgl. 1996) gezeigt). Viele delegieren solche Entscheidungen anstatt fair zu teilen, wenn sie sich dadurch besser stellen können (vgl. Andreoni/Bernheim 2009; Hamman et al. 2010; Bartling/Fischbacher 2012). Das Kaschieren eigeninteressierter Verhaltenstendenzen ist möglich, soweit plausible Ausreden bzw. entsprechend alternative Erklärungen zur Verfügung stehen. Dana et al. (vgl. 2007) führten z.B. die Regel ein, dass nach spätestens 10 Sekunden Wartezeit ein Computer eine Zufallsentscheidung trifft, bei der zu 50 Prozent die „faire“ Allokation hergestellt und zu 50 Prozent zum Vorteil des Diktators verteilt wird (Plausible-Deniability-Bedingung).<sup>4</sup>

Während im Baseline-Treatment, bei dem der Computer erst nach 60 Sekunden intervenieren würde, 74 Prozent der Diktatoren die faire Variante wählen, wählen unter der Plausible-Deniability-Bedingung von denen, die nicht tatsächlich vom Computer abgeschnitten wurden, nur 45 Prozent die faire Variante. 55 Prozent handelten entsprechend eigeninteressiert. Darüber hinaus wurden 24 Prozent der Diktatoren tatsächlich abgeschnitten. Dies ist bemerkenswert, weil in einem Pretest ermittelt wurde, dass niemand länger als 2 Sekunden benötigt, um sich zu entscheiden, und der Computer tatsächlich frühestens nach 4 Sekunden intervenierte.

Diese Befunde führten zu der Überzeugung, dass man das positive, von Fairness geprägte Menschenbild ebenfalls revidieren müsse, weil die Befunde mit der Hypothese einer grundlegenden intrinsischen „Ungleichheitsaversion“, wie sie von Fehr und Schmidt (vgl. 1999) sowie Bolton und Ockenfels (vgl. 2000) unterstellt worden war, unvereinbar sind. Dana et al. bringen das so auf den Punkt: „Rather than having a preference for a fair outcome, people may conform to situational pressures to give in certain contexts, but may also try to exploit situational justifications for behaving selfishly“ (2007: 69).

Das bedeutet nun aber nichts anderes, als dass man (bei den meisten Menschen) eben nicht von grundlegenden moralischen Motiven ausgehen dürfte, sondern eine Orientierung am Eigeninteresse unterstellen müsste, und dass die scheinbare Moral danach nur durch den drohenden Entzug sozialer Anerkennung und damit über Handlungsanreize induziert werde. Der Homo Oeconomicus ist tot – es lebe der Homo Oeconomicus!?

---

<sup>4</sup> Im Experiment von Dana et. al gibt es genau zwei Möglichkeiten (A und B). Im Fall A erhält der Diktator \$6, der Rezipient \$1. Im Fall B erhalten beide \$5. A ist daher nicht nur tendenziell unfair, sondern auch weniger effizient als B, weil die Gesamtauszahlung hier nur \$7 beträgt.

Dies wäre sicherlich die falsche Konsequenz, die vermutlich auch niemand zieht. Dagegen spricht neben einer allzu simplen „Pendelschlag-Logik“<sup>5</sup> schon der Umstand, dass moralischer Druck nur dann ausgeübt werden kann, wenn es zumindest eine hinreichende Anzahl entsprechend „moralischer“ Individuen gibt, die diesen Druck aufbauen und aufrechterhalten können. Gleichwohl steht jedoch die Frage im Raum, ob denjenigen, die den Anreizen zu eigeninteressiertem Handeln nachgeben, eine intrinsische Moralität bzw. moralische Motivation abzusprechen ist. Auch eine solche Argumentation wäre m.E. kurzschlüssig.

Zumindest ist es möglich, der Hypothese, dass viele nur „moralisch“ handelten, um sozial anerkannt zu werden, eine Alternativhypothese entgegenzusetzen. Statt zu sagen, Individuen seien grundlegend und stabil eigeninteressiert und verhielten sich in bestimmten Situationen nur so, als ob sie moralisch handelten, lässt sich ebenso vermuten, sie verfügten über grundlegende und stabile moralische Präferenzen und verhielten sich nur unter entsprechenden situativen Anreizstrukturen wie eigeninteressierte Akteure (vgl. hierzu auch Bénabou/Tirole 2011).

### 2.3 Framing-Effekte und Rationalität

Zu dieser Alternativhypothese geben nicht nur wirtschaftsethische Arbeiten Anlass (vgl. Homann 2004; Pies 2009; Pies/Hielscher 2014; Homann 2014)<sup>6</sup>, sondern auch neuere Erkenntnisse zu Framing-Effekten. Diese sind im vorliegenden Kontext besonders interessant, weil sie einerseits zeigen, dass Framing eher an den Restriktionen als an den Präferenzen ansetzt, und andererseits, dass differentielles Verhalten in Frames durchaus rational sein kann.

Framing-Effekte wurden bereits seit Anfang der 1980er Jahre, vor allem von Kahneman und Tversky, erforscht (vgl. Kahneman 2012: 363–374). Kahneman spricht dabei von inkonsistenten Präferenzen und unbegründeten Einflüssen von Frames auf Entscheidungen (vgl. auch Tversky/Kahneman 1986). Insbesondere unterscheidet Kahneman in diesem Zusammenhang zwischen zwei Systemen der menschlichen Handlungssteuerung: einem intuitiven und schnell operierenden „System 1“ und einem reflektierten, aber langsamen „System 2“. Framing-Effekte beeinflussen danach insbesondere System 1, und Entscheidungen im Rahmen dieses Systems können, so die herrschende Auffassung, irrational sein.<sup>7</sup> Die Annahme der Irrationalität von System-1-Entscheidungen ist aber problematisch.

---

<sup>5</sup> Damit ist gemeint, dass der Mensch abwechselnd mal in seiner Moralität „gefeiert“, mal in seiner augenscheinlichen Unmoral „geächtet“ wird.

<sup>6</sup> Nickamp (vgl. 2014) weist allerdings zu Recht darauf hin, dass unter dieser Hypothese im ordnungsethischen Rahmen Differenzierungen in der Analyse von Person- und Situationsmerkmalen sowie der Interaktion zwischen Person und Situation vorzunehmen sind. Der vorliegende Beitrag versteht sich als ein Ansatz dazu, diesem Anspruch gerecht zu werden.

<sup>7</sup> Man denke bspw. an das bekannte Asian-Disease-Problem, bei dem es um eine Krankheit geht, die das Leben von 600 Menschen bedroht. Frame 1 („Leben retten“): Man hat die Wahl zwischen einem Programm, bei dem 200 Menschen gerettet werden, und einem Programm, bei dem mit einer Wahrscheinlichkeit von 1/3 alle gerettet werden und ansonsten niemand gerettet wird. Frame 2 („Sterben“): Man wählt zwischen einem Programm, bei dem 400 Menschen sterben werden, und einem Programm, bei dem mit einer Wahrscheinlichkeit von 1/3 niemand sterben

*Erstens* ist intuitives Handeln oftmals sehr intelligentes Handeln (vgl. z.B. Neuweg 2015; Smith 2008), und die Prozesse, auf denen es beruht, können als implizite inferentielle Prozesse rekonstruiert werden (vgl. Minnameier 2016). Die vermutete Irrationalität resultiert offenbar nur daraus, dass man intuitives Denken und Handeln nach den Kriterien von System 2 beurteilt, anstatt die innere Logik von System-1-Prozessen zu untersuchen.<sup>8</sup> *Zweitens* konnte gerade für soziale Präferenzen gezeigt werden, dass sich das – vermutlich weitgehend intuitiv gesteuerte – Verhalten in Experimenten sehr wohl mit dem Rational Choice-Ansatz verträgt (vgl. Gächter 2013). *Drittens* haben Ellingsen et al. (2012) herausgearbeitet und empirisch belegt, dass Framing-Effekte nicht auf framespezifische Präferenzen zurückzuführen sind, sondern von den jeweils wahrgenommenen Restriktionen abhängen. Hierauf soll im Folgenden näher eingegangen werden.

Die Untersuchung von Ellingsen et al. (ebd.) setzt an den Erkenntnissen zum Framing im Kontext des Gefangenendilemmas an. Wie vor allem Liberman et al. (vgl. 2004) eindrucksvoll demonstriert haben, hat allein die Namensgebung einen erheblich Einfluss auf das Spielverhalten. Sie nannten das Gefangenendilemma in einen Fall „Wall Street Game“ im anderen Fall „Community Game“. Ansonsten war alles gleich. Bei ihrer Untersuchung handelte es sich allerdings um die Variante mit Wiederholungen, bei der zwei identische Spielpartner acht Runden miteinander spielten. Nach diesen acht Runden kooperierte im „Wall Street Game“ etwa ein Drittel<sup>9</sup>, im „Community Game“ waren es rund zwei Drittel.

Ellingsen et al. (vgl. 2012) ermittelten darüber hinaus, dass sich bei einem One-Shot-Game ein ähnliches Muster ergibt. Obwohl das Spiel unter beiden Frames ein simples Gefangenendilemma ist, bei dem die Teilnehmer sowohl die Auszahlungen je Entscheidungsalternative als auch die komplette Auszahlungsmatrix präsentiert bekommen haben, ist die Kooperationsrate im „Community Game“ viel höher als im „Stockmarket Game“.

Der Name des Spiels macht also schon den entscheidenden Unterschied. Allerdings verschwindet dieser Framing-Effekt fast vollständig (und ist nicht mehr signifikant), wenn man das One-Shot-Game sequentiell spielen lässt und die Respondenten damit auf einer gesicherten Grundlage entscheiden können. Knapp 40 Prozent der Spieler 1 (in der Rolle der Proposer) kooperierten, unabhängig vom Frame und obwohl ihre finanzielle Auszahlung beim Defektieren höher gewesen wäre. Außerdem konnten sie ja nicht wissen, wie die Respondenten entscheiden würden. Wenn Spieler 1 kooperierte,

---

wird, ansonsten alle sterben werden. Beide Frames haben das gleiche Resultat, aber im ersten Frame wird häufiger die sichere Lösung gewählt als im zweiten Frame (vgl. Kahneman 2012; Tversky/Kahneman 1986). Das Verhalten wird demnach offenbar von irrelevanten Aspekten bestimmt.

<sup>8</sup> Beim Asian-Disease Problem (vgl. Fußnote 7) könnte es eine Rolle spielen, dass man im ersten Frame von 200 mit Sicherheit Überlebenden ausgeht, was jedoch nicht ausschließt, dass es noch mehr sein könnten. Wenn Gleiches für die Sterbenden im zweiten Frame gilt, dann würde man hier mit mindestens 400 Toten rechnen, und damit wäre ein relevanter Unterschied zwischen beiden Situationen konstituiert.

<sup>9</sup> Dass es hier mehr sind als die 10 Prozent Kooperierer im One-Shot-Gefangenendilemma (vgl. Ledyard 1995), erklärt sich dadurch, dass bei Spielwiederholung eine Tit-for-Tat-Strategie möglich ist, über die wechselseitige Kooperation aufgebaut werden kann (vgl. Axelrod 1984/2005).

kooperierten auch ca. 45 Prozent der Spieler 2, wiederum unabhängig vom jeweiligen Frame. Es lässt sich also festhalten, dass insgesamt fast 50 Prozent der Teilnehmer lieber kooperieren, als den finanziellen Vorteil zu suchen, wenn sie wissen oder zumindest erwarten können, dass der jeweilige Mitspieler kooperiert. Angesichts der Tatsache, dass es sich um ein Spiel zwischen einander nicht bekannten Akteuren unter einer Anonymitätsbedingung handelt, ist dies sicherlich als eine recht hohe Kooperationsrate einzuschätzen. Man könnte ja auch argumentieren, dass es in einer solchen Situation moralisch völlig in Ordnung sei, dass jeder Akteur seinem Eigeninteresse folgt (eine Haltung, die in Untersuchungen zum Diktatorspiel auch belegt ist).

Offenbar entscheiden die Akteure rational im Sinne ihrer jeweiligen Präferenzen (die freilich interindividuell verschieden sind) und der wahrgenommenen Restriktionen. Framing wäre nur dann irrational, wenn die Handelnden intraindividuell in identischen Situationen jeweils verschiedene Präferenzen hätten. Genau das ist aber nach bisherigem Kenntnisstand nicht der Fall.

Indizien für eine solche Anpassung an (wahrgenommene Restriktionen) liefern außer der Gruppe um Ellingsen auch Cohn et al. (vgl. 2014), die anhand verschiedener Experimente zeigen konnten, dass Bankangestellte besonders anfällig für unaufrichtiges Verhalten sind, das allerdings nur dann, wenn Aspekte ihrer professionellen Identität als „Banker“ durch Priming hervorgehoben werden. In einem Kontrollexperiment verhielten sie sich im Allgemeinen ehrlich. Darüber hinaus konnten sie zeigen, dass dieser Effekt im Wesentlichen von ihrer Einschätzung dessen abhängt, was man von ihnen erwartet und wofür sie Anerkennung erfahren. Je mehr man glaubt, finanzieller Erfolg sei besonders wichtig, desto größer ist die Neigung im Experiment zu mogeln. Ausschlaggebend sind also offenbar auch hier spezifische Restriktionen, nicht grundlegende moralische Präferenzen.

## 2.4 Weiter Vorteilsbegriff und soziale Präferenzen

Ellingsen et al. (vgl. 2012) verweisen im Zusammenhang stabiler Präferenzen auf Becker und Stigler (vgl. 1977), auf die die Annahme fundamentaler und stabiler Präferenzen im Wesentlichen zurückgeht (vgl. auch Dreber et al. 2013). Im Rahmen dieses Ansatzes werden allerdings Präferenzen ganz anders gefasst als z.B. im bis dahin führenden und auch heute noch verwendeten Ansatz der „bekundeten Präferenzen“ (revealed preferences) (vgl. Samuelson 1948; vgl. auch Hollis/Sugden 1993; Binmore 2009). Während bekundete Präferenzen auf konkrete Entscheidungsobjekte – typischerweise Güterbündel – referieren, beziehen sich Präferenzen im Becker’schen Ansatz prinzipiell auf alles, was Menschen explizit oder implizit als wertvoll erachten, mithin was ihren konkreten Entscheidungen für bestimmte Entscheidungsobjekte unter bestimmten Restriktionen zugrunde liegt (siehe auch Abschnitt 2.1).

Dieser weite Vorteilsbegriff wurde auch in der Wirtschaftsethik aufgegriffen (vgl. z.B. Homann 1994; Homann/Suchanek 2005; Homann 2014). Weitgehend offen geblieben ist allerdings, wie dieser weite Vorteilsbegriff inhaltlich gefasst werden könnte, insbesondere im Hinblick auf moralische Präferenzen. Bezüglich der Becker’schen Modellierung liegt das vor allem daran, dass „Moral“ dort gar nicht zu den Präferenzen gehört, sondern zu den Restriktionen. Präferenzen im Becker’schen Sinne sind nämlich noch grundlegender und beziehen sich auf basale Motive, die im Großen und Ganzen für alle



Menschen die gleichen sind. Sie sind deshalb „stable over time and similar among people“ (Becker/Stigler 1977: 76). Moral zählt für Becker insofern zu den Restriktionen, als sie zum persönlichen Human- und Sozialkapital gehört, die wiederum Determinanten in der Becker'schen Haushaltsfunktion sind. Das bedeutet, dass Nutzen im Sinne der Grundbedürfnisse unter Einsatz dieser Kapitalien produziert werden muss, und diese Kapitalien folglich Möglichkeiten für die Nutzenproduktion schaffen oder beschränken (vgl. Becker 1996; vgl. auch Vanberg 2008).

So kann man natürlich im Rahmen ökonomischer Verhaltenserklärungen ansetzen, verfehlt damit aber die Perspektive der individuellen Handlungssteuerung, die ordnungsethisch und institutionenökonomisch von zentraler Bedeutung ist (vgl. Niekamp 2014). Denn aus der Perspektive des Subjekts sind moralische Überzeugungen keine Restriktionen, sondern grundlegende Ziele und damit Handlungsmotive, sprich: Präferenzen. Will man diese „Präferenzen“ angemessen berücksichtigen, muss man mindestens drei Aspekte unterscheiden:

1. die für alle Menschen gleichen oder ähnlichen Grundbedürfnisse,
2. substantielle, interindividuell verschiedene, aber intraindividuell weitgehend stabile Präferenzen (die sich allenfalls entwicklungsbedingt, aber nicht situationsbedingt verändern) und
3. intra- und interindividuell variable konkrete Entscheidungen, wobei die intra-individuelle Variabilität der Letzteren von der Variabilität der (wahrgenommenen) Restriktionen abhängt und durch diese erklärt wird (wie im obigen Beispiel zum Framing).

Die Differenzierung von (2.) und (3.) ist vor allem wichtig, wenn Akteure – wie etwa im Homann'schen Ethikansatz – systematisch zwischen Handlungen und Handlungsbedingungen unterscheiden sollen. Motiviert werden Handlungen durch explizite moralische Prinzipien oder implizite Handlungstendenzen (psychische Dispositionen) im Sinne von (2.). Die Handlungsbedingungen werden im Rahmen von (3.) verarbeitet.

Die in der Verhaltensökonomik diskutierten „sozialen Präferenzen“ wie z.B. „Inequity Aversion“ (vgl. Fehr/Schmidt 1999; Bolton/Ockenfels 2000) sind prinzipiell Kandidaten für die gemäß (2.) gesuchten substantiellen Präferenzen. Allerdings weisen die hierzu vorliegenden Ansätze m.E. zwei grundlegende Schwächen auf: *Zum einen* dürften die bisherigen Konzeptualisierungen den Reichtum an moralischen Prinzipien, denen sich Menschen verpflichtet fühlen können, nicht umfänglich erfassen. Es handelt sich um eher einfache Prinzipien wie die schon genannte Ungleichheitsaversion oder das einfache Reziprozitätsprinzip (Gleiches mit Gleichem zu vergelten). Konzepte wie „Intention-based Reciprocity“ (vgl. Rabin 1993; Dufwenberg/Kirchsteiger 2004; Falk/Fischbacher 2006; Falk et al. 2008) greifen zwar darüber hinaus, aber die Verpflichtung auf eine soziale Einheit (wie z.B. für den Erfolg und das Gedeihen eines Unternehmens) oder auf kulturell geprägte Werte als normative Basis (z.B. ob ein Handschlag gelten muss oder nicht) werden dadurch nicht erfasst – ganz zu schweigen von ethischen Prinzipien im engeren Sinn.

*Zum anderen* – und dies wiegt im vorliegenden Zusammenhang schwerer – wird auch im Kontext sozialer Präferenzen nicht systematisch zwischen Präferenzen und Restriktionen differenziert, jedenfalls nicht in dem Sinne, dass bestimmte Präferenzen unter bestimmten Restriktionen zum Tragen kommen können, in anderen dagegen nicht. Zwar zeigen die Ergebnisse verhaltensökonomischer Experimente, dass z.B. die Kooperation prinzipiell ungleichheitsaverser Akteure in bestimmten Situationen abnimmt (vgl. List 2007; Dana et al. 2007) oder dass je nach Situation verschiedene präferenzbezogene Orientierungen zu verzeichnen sind (vgl. Engelmann/Strobel 2004). Aber es wird nicht systematisch untersucht, warum diese jeweiligen Präferenzen in bestimmten Situationen bevorzugt gewählt werden, in denen sie eine jeweils spezifische Relevanz zu besitzen scheinen. Stattdessen fragen Fehr und Schmidt (2006: 668):

„Which model of other-regarding preferences does best in the light of the data, and which should be used in applications to economically important phenomena? We believe that it is too early to give a conclusive answer to these questions. There is a large amount of heterogeneity at the individual level and any model has difficulties in explaining the full diversity of the experimental observations.“

Meines Erachtens wäre jedoch davon auszugehen, dass *alle* Modelle sozialer Präferenzen eine *spezifische* Relevanz haben und in bestimmten Kontexten handlungsrelevant sind, in anderen dagegen nicht. Der Aspekt der situationsspezifischen Relevanz und Nutzung moralischer Prinzipien bleibt somit bislang eher ausgeblendet; jedenfalls wird er nicht systematisch im Sinne einer Situationstheorie moralischen Handelns aufgegriffen. Dabei gibt es seit langem und immer wieder empirische Hinweise darauf (und theoretische Ansätze dazu), dass Menschen nicht nur über unterschiedliche Moralprinzipien verfügen, sondern diese auch situationsspezifisch einsetzen (vgl. Beck et al. 2002; Beck/Parche-Kawik 2004; Krebs/Denton 2005; Lindenberg 2008; Binmore/Shaked 2010; Minnameier/Schmidt 2013; Kaplan/Tivnan 2014).

Dass die Modelle der Verhaltensökonomik in diesem Punkt bislang wenig elaboriert sind, hat auch historische Gründe. In einer über zwei bis drei Jahrhunderte langen Entwicklung wurde der Nutzenbegriff inhaltlich völlig entleert (vgl. Hollis/Sugden 1993; Bruni/Sugden 2007). In der neoklassischen Theorietradition ging man schließlich von einer grundlegenden Subjektivität der Werte aus und setzte daher am subjektiven Nutzen an, der jedoch als solcher schwer zu messen war und zum paretianischen ordinalen Nutzenbegriff führte. Der letzte Schritt bestand in Samuelsons Konzept bekundeter Präferenzen, wonach Präferenzen (die Nutzen stiften) nur noch von Entscheidungen abgeleitet wurden (wo eigentlich umgekehrt Entscheidungen von Präferenzen und Restriktionen abzuleiten wären). „The notion of utility has thus been bleached of all psychological content and we are left with an abstract scheme of quasi-logical relations“ (Hollis/Sugden 2007: 7).

Offen bleiben bei diesem Stand der Diskussion vor allem zwei Fragen, denen der folgende Abschnitt (insbesondere 3.2 und 3.3) gewidmet ist:

1. Welche moralischen Präferenzen lassen sich unterscheiden, und wie bauen sie logisch aufeinander auf?

2. Wie funktioniert die situationsspezifische Aktivierung, und wie lässt sich diese im Rahmen des RCT erklären.
3. „Reason-Based Theory“ (RBT) und Moralität

### 3.1 Gründe und die Bildung von Präferenzen

Vor nicht allzu langer Zeit haben Dietrich und List (vgl. 2013a; 2013b) eine „Reason-Based Theory of Rational Choice“ (RBT) aufgestellt, auf deren Basis die oben herausgearbeiteten Probleme gelöst werden könnten. In der RBT nehmen Gründe eine zentrale Rolle ein, und zwar insbesondere als die (eher) stabilen Persönlichkeitsmerkmale, von denen (konkrete) Präferenzen ihrerseits abhängen. „Präferenzen“ selbst werden dabei als situationsspezifisch variabel aufgefasst, was bedeutet, dass dieser Begriff von Präferenzen weder mit dem Becker’schen noch mit dem Ellingsen’schen Begriff kompatibel ist. Allerdings geht die Idee „fundamentaler“ Präferenzen in den „Gründen“ auf, so dass man, wie die Autoren selbst bemerken, auch von zwei Arten von Präferenzen – „fundamentalen“ und „abgeleiteten“ – sprechen könnte. Zugleich argumentieren sie gegen die überkommene Vorstellung (die auch bei Becker präsent ist), fundamentale Präferenzen seien nicht-kognitive Strebungen (desires):

„A rational agent, on the standard picture, has beliefs and desires (typically modelled as assignments of probabilities and utilities to different possible worlds, states or outcomes), and acts so as to satisfy his or her desires in accordance with his or her beliefs. On this picture, the agent’s desires over possible worlds or fully specified outcomes—his or her *fundamental preferences*—are volitional attitudes that are entirely separate from, and do not respond to, the agent’s beliefs, which are cognitive attitudes. Only desires over prospects with uncertain outcomes can change in response to belief changes about the relative likelihood of these outcomes; these ‘surface-level’ desires are also called *derived preferences*.“ (Dietrich/List 2013a: 104)

Trotz dieses Konzepts von (abgeleiteten) Präferenzen unterscheiden die Autoren zwischen der Bildung von Präferenzen einerseits und dem Handeln (Choice) andererseits. Dieser Aspekt wird in Abschnitt 3.4 aufgegriffen.

„Gründe“ verstehen Dietrich und List als besondere Arten von Propositionen, auf die die Präferenzbildung zurückgeführt werden kann. Sie unterscheiden dabei zwischen „motivierenden“ und „normativen“ Gründen (ebd.: 106f.). Erstere begründen, warum bestimmte Präferenzen vorliegen; Letztere begründen, welche Präferenzen man (aus Sicht der urteilenden Person) haben sollte. In ihren Beispielen diskutieren Dietrich und List Fälle alltäglicher Entscheidungen und beschränken sich dabei auf motivierende Gründe. Es sei jedoch hinzugefügt, dass bei moralischen Entscheidungen gerade moralische Prinzipien als „motivierende Gründe“ ins Spiel kommen – als Gründe, eine moralische Entscheidung so oder anders zu treffen.

Der Ansatz geht aus von bestimmten Gegenständen („Alternativen“ genannt), über die der Akteur Präferenzen bilden muss. Grundlegend ist also eine beliebige Entscheidungssituation. Die Menge der Alternativen wird mit  $X$  bezeichnet, und Präferenzen werden durch eine Ordnung  $\succeq$  auf  $X$  ausgedrückt. Für beliebige Alternativen  $x$  und  $y$

bedeutet dann  $x \succeq y$ , dass  $x$  gegenüber  $y$  schwach präferiert wird.  $x \succ y$  bezeichnet eine starke Präferenz und  $x \sim y$  Indifferenz.

Die Alternativen könnten beispielsweise zwei Autos sein, zwischen denen sich ein potentieller Käufer entscheiden muss, sagen wir zwischen einem Audi Q7 und einem VW Golf. Abgesehen vom unterschiedlichen Preis spielen sicherlich weitere Faktoren eine Rolle für die Entscheidung: z.B. Größe (G), Schnelligkeit (S) und Umweltfreundlichkeit (U). Alle diese Faktoren kommen nun als motivierende Gründe in Frage, und welches Fahrzeug am Ende präferiert wird, hängt von den Gründen und ihrer relativen Bedeutung für den Akteur ab. Die durch motivierende Gründe induzierte Präferenz wird mit  $x \succeq_M y$  bezeichnet, wobei  $M$  für die betreffende Menge motivierender Gründe steht.

Um die relative Bedeutung von Gründen erfassen zu können, wird zudem eine Gewichtungsrelation benötigt. Wenn z.B. die einfache Rangordnung  $U > G > S$  gilt, so wird der Golf präferiert, weil er umweltfreundlicher ist als der Q7. Würde Umweltfreundlichkeit in der Situation des Autokaufs jedoch keine Rolle spielen, so würde der größere Q7 präferiert.

Eine noch offene Frage ist allerdings, wie die Gewichtungsrelation zu interpretieren ist. Dietrich und List (2013a: 113) sehen hier die zwei Möglichkeiten einer kognitiven und einer nicht-kognitiven Interpretation. Nach der Ersteren würden Urteile zugrunde liegen, aus denen die relative Wichtigkeit der jeweiligen Gründe hervorgeht, nach der Letzteren enkodiert die Gewichtungsrelation schlicht die persönlichen Dispositionen, die den Akteur zu seiner Gewichtung führen. Dietrich und List lassen offen, welche Interpretation vorzuziehen wäre (ebd.).

Aus meiner Sicht ist tatsächlich beides denkbar, insbesondere wenn man zwischen moralischen und ästhetischen Fragen unterscheidet. Geht es nur um die Frage des persönlichen Geschmacks (ästhetische Frage), wäre die nichtkognitive Interpretation zumindest denkbar, denn dann könnte man womöglich nicht viel mehr sagen, als dass einem eben der eine Aspekt wichtiger ist als ein anderer. Geht es jedoch um Moral, so *müsste* man sogar Gründe haben, warum man z.B. ein umweltfreundliches Auto vorzieht, denn diese Gründe sollen ja auch *andere* überzeugen können (vgl. Birnbacher 2007).

Moralische Urteile über die Vorzugswürdigkeit bestimmter relevanter Eigenschaften müssen insofern auf übergeordnete Gründe Bezug nehmen. (Dieser Punkt ist bei Dietrich und List nicht weiter ausgeführt, weil sie sich auch nicht speziell mit moralischen Fragen beschäftigen.) Darüber hinaus ist im Bereich der Moral sogar von einer hierarchischen Struktur von Begründungen bzw. Begründungsprinzipien auszugehen. Das Gebot der Toleranz bezieht sich z.B. auf die Unterschiedlichkeit moralischer Beurteilungen durch verschiedene Akteure bzw. Kulturen und hat insofern moralische Beurteilungen zum *Inhalt*, die zugleich ein moralisches Problem höherer Ordnung konstituieren.

Dieser hierarchische Aspekt ist auch relevant für die Frage der Stabilität „fundamentaler Präferenzen“. Wenn Gründe „fundamentale Präferenzen“ sind, dann gilt das nämlich stets (nur) relativ zu den auf ihrer Basis gebildeten Präferenzen, denn nur für sie sind sie „fundamental“.

### 3.2 Fundamentale Präferenzen im Sinne moralischer Prinzipien

Die in Abschnitt Framing-Effekte und Rationalität 2.3 angesprochene Neuinterpretation von Framing-Effekten, bei der strikt und stringent zwischen stabilen (fundamentalen) Präferenzen einerseits und wahrgenommenen Restriktionen (Beliefs) andererseits unterschieden wird, lässt sich auf Basis der RBT weiter elaborieren. Soweit bei denjenigen, die im Gefangenendilemma kooperieren, moralische Überlegungen den Ausschlag geben, könnte das zugrunde liegende Prinzip das der sog. Goldene Regel sein: „Was du nicht willst, das man dir tu“, das füg’ auch keinem andern zu!“ Darin kommt der eben schon erwähnte Anspruch zum Ausdruck, auch auf die Belange der anderen Beteiligten Rücksicht zu nehmen und so von einem übergreifenden Standpunkt aus zu entscheiden.

Die Goldene Regel wäre hier die fundamentale Präferenz. Die Entscheidung zu kooperieren wäre entsprechend die davon abgeleitete Präferenz im Sinne Dietrichs und Lists. Über die Spielstruktur als solche hinaus kommen jedoch zusätzliche Restriktionen ins Spiel, wenn man bedenkt, dass die Goldene Regel vernünftigerweise nur dann anwendbar ist, wenn das eigene moralische Verhalten nicht durch andere ausgebeutet wird. Im Gefangenendilemma ist dies das Grundproblem, das beim Stockmarket-Framing nicht nur besonders salient, sondern wie gezeigt auch besonders virulent ist. Je nach Potentialen und Beschränkungen, die in einer bestimmten Situation gesehen werden, könnten entsprechende Prinzipien aus dem Repertoire individuell verfügbarer Moralkonzepte aktiviert und ausgewählt werden. Spezifische moralische Probleme führen so zu spezifischen moralischen Lösungen. Wie die Verarbeitung solcher situationsspezifischer Restriktionen verstanden werden kann, wird in Abschnitt 3.3 ausgeführt. Hier wird im Folgenden zunächst eine Taxonomie vorgestellt, die die hierarchische Struktur moralischer Prinzipien erklärt.

Moralische Prinzipien sind vielfältig. Wo unser Mitgefühl angesprochen ist, werden altruistische Tendenzen wirksam. In Wettbewerbssituationen gilt hingegen, dass sich jeder – freilich unter Einhaltung der Spielregeln – selbst der Nächste ist (weil er sonst ausgebeutet wird und eine solche Ausbeutung nicht hinzunehmen ist). Ist strategische Kooperation jedoch möglich, so greift eher das Prinzip „Wie du mir, so ich dir“, oder das der expliziten Aushandlung von Verträgen (z.B. durch Spielen einer Tit-for-Tat-Strategie). Es gibt aber auch Kontexte, in denen statt konkreter Verträge die gemeinsame Verantwortung für die Organisation bzw. die soziale Einheit, der man sich zugehörig fühlt, im Mittelpunkt steht, beispielsweise für die eigene Familie oder das Unternehmen, bei dem man beschäftigt ist. Eine solche „soziale“ Verantwortung kann aber wiederum nur greifen, wenn das Engagement des Einzelnen nicht von anderen strategisch ausgenutzt wird. Deshalb muss es Sicherungsmaßnahmen gegen die Möglichkeit des Trittbrettfahrens geben. Man sieht schon daran, dass die Prinzipien aufeinander aufbauen und höhere Prinzipien die Erfüllung der niedrigeren implizieren bzw. voraussetzen.

Die Stufen selbst lassen sich z.B. mit Hilfe einer neo-Kohlberg’schen Stufentaxonomie explizieren, die an der Kohlberg-Theorie (vgl. z.B. Colby/Kohlberg 1987; Kohlberg 1997) ansetzt, diese aber in verschiedenerlei Hinsicht weiterentwickelt. Abgesehen davon dass die neo-Kohlberg’sche Taxonomie mehr und differenziertere Stufen aufweist als die Kohlberg’sche, basiert die Erstere auf einer Entwicklungslogik, die bei Kohlberg zwar ebenfalls behauptet, aber nie expliziert wurde (vgl. hierzu Minnameier 2000; 2010a).

Die neo-Kohlberg'sche Taxonomie basiert auf einer Entwicklungslogik die Piaget in seinem Spätwerk entwickelt hat und darin besteht, dass ein Gegenstand zunächst aus unterschiedlichen Perspektiven betrachtet wird („intra“), dann zwischen diesen Perspektiven Verbindungen hergestellt („inter“) und diese schließlich zu einer vereinheitlichten Perspektive zusammengeführt werden („trans“) (vgl. Piaget/Garcia 1989). Übertragen auf den Bereich der Moral bedeutet dies, dass zunächst individuelle Perspektiven differenziert werden (z.B. „mein Eigentum, dein Eigentum“), dann diese Perspektiven reziprok miteinander verknüpft werden (z.B. „Ich gebe Dir etwas ab, und du gibst mir etwas ab“) und schließlich aus einer übergreifenden Perspektive integriert werden (z.B. „Ich habe viel, und du hast wenig, also gebe ich dir etwas ab, dann ist es ausgeglichen“). In Anlehnung an Kohlberg (vgl. Kohlberg et al. 1984) spreche ich hier von „Gerechtigkeitsoperationen“ und nenne diese „Gleichheit“, „Reziprozität“ und „reziproke Gleichheit“. Sie lassen sich folgendermaßen definieren und erläutern:

1. *Gleichheit*: Die Gleichheitsoperation bedeutet, dass individuelle Perspektiven eingenommen werden können und als prinzipiell gleichberechtigt nebeneinander gestellt werden. Im einfachsten Fall beansprucht man das eigene Eigentum und respektiert das der anderen. Aber es gilt auch für den Grundsatz, jeder Mensch habe Interessen und soll diesen gemäß rational handeln, oder auch für das Prinzip, jeder müsse nach seinem persönlichen Gewissen entscheiden, oder für Gleichberechtigung kulturrelativer Ethiken. In allen Fällen werden Positionen als gleichberechtigt anerkannt und nebeneinander gestellt, es kann aber nicht zwischen ihnen vermittelt werden.
2. *Reziprozität*: Durch die Reziprozitätsrelation wird zwischen Perspektiven vermittelt, die zuvor im Sinne der Gleichheitsoperation (1.) differenziert werden. Das geschieht etwa durch das Prinzip des gleichen Teilens eines gemeinsamen Gutes (wie etwa im Diktatorspiel), aber auch durch Verträge, in denen man Leistung und Gegenleistung vereinbart und damit einen Ausgleich zwischen subjektiven Interessen herstellt (also nicht objektiv vergleichbare Wertigkeiten durch Aushandlung zum Ausgleich bringt). Andere Formen von Reziprozität liegen vor, wenn man die Problematik einer reinen Gewissensethik zu überwinden versucht, etwa indem man sagt, es müsse ein gleiches Recht für alle gelten bzw. einen „Schiedsrichter“, dessen Urteil von allen akzeptiert werden müsse; oder im Bereich der Metaethik, wo z. B. der Hobbes'sche Sozialvertrag die reziproke Antwort auf einen strikt subjektbezogenen Rationalismus darstellt.
3. *Reziproke Gleichheit*: Bei der Reziprozität werden die Unterschiede zwischen den zunächst differenzierten Perspektiven ausgeblendet und nur ein Entscheidungsmechanismus zwischen die Perspektiven eingebaut. Diese Unterschiede werden aber z.B. bedeutsam, wenn ungleiche Akteure etwas teilen sollen (soll ein armer Diktator einem reichen Rezipienten etwas abgeben, und wenn ja: wie viel?). Oder wenn ein Vertragspartner eine besondere Machtstellung hat, besteht die Gefahr des Wuchers, so dass reziproke Aushandlung nicht mehr zu einem fairen Ergebnis führt. Oder wenn der o.g. Schiedsrichter in seinem Urteil angezweifelt wird, dann benötigt man ein Prinzip, wie man gerechte allgemeine

Regeln gewinnt. Das Vertragsproblem lässt sich durch die einfache goldene Regel lösen, das Regelungsproblem durch demokratische Verfahren (wenn man davon ausgeht, dass sich dabei jeder gleichberechtigt einbringen kann). Ebenso lässt sich für das o.g. metaethische Beispiel sagen, dass Hobbes kein Kriterium dafür hatte, verschiedene (Hobbessche) Sozialverträge im Hinblick auf ihre Vorzugswürdigkeit zu unterscheiden; genau das beansprucht darüber hinausgehend z.B. Rousseau, indem er einen Sozialvertrag konzipiert, den ein Vernunftmensch mit sich selbst schließt und der die „Volonté générale“ (im Unterschied zu einer „Volonté de tous“) zum Ausdruck bringt.

Insgesamt ergibt sich eine Taxonomie der folgenden Struktur (vgl. Abbildung 1), die hier freilich nicht im Einzelnen expliziert werden kann (vgl. hierzu Minnameier 2000; 2005; 2010a). Es wird jedoch deutlich, dass „Gleichheit“ und „Reziprozität“ in verschiedenen moralrelevanten Zusammenhängen eine Rolle spielen, dabei aber auch eine jeweils verschiedene Bedeutung haben. In diesem Rahmen lassen sich prinzipiell auch die in der Verhaltensökonomik diskutierten sozialen Präferenzen verorten. Allerdings wären dazu entsprechend differenzierte Begriffsanalysen erforderlich, denn die Bedeutungen „sozialer Präferenzen“ sind in aller Regel lediglich extensional bestimmt und daher auf experimentelle Settings beschränkt (vgl. zu diesem Problem auch Levitt/List 2007).

Ebenen	I				II				III			
Stufen	1	2	3	4	5	6	7	8	9			
Unter- stufen	A	B	C	A	B	C	A	B	C	A	B	C

*Abbildung. 1: Die neo-Kohlberg'sche Stufentaxonomie im Überblick  
(Quelle: eigene Darstellung)*

Der zentrale Unterschied zwischen Stufe 1 und Stufe 2 besteht darin, dass Stufe 1 auf Mitgefühl im Rahmen sozialer Perspektivenübernahme basiert, während Stufe 2 jenseits des unmittelbaren Mitgefühls individuelle „Interessen“ koordiniert, die einander gerade entgegenstehen (vgl. Abbildung 2). Ego kann sich auf Stufe 1 in Alter hineinversetzen, und weil Ego das tut und mit Alter mitfühlt und mitleidet, ist Ego motiviert zwischen beiden Perspektiven einen Ausgleich herzustellen. Ebenso versteht man auf Stufe 1 grundlegende moralische Regeln, eben weil man die Perspektive anderer einnehmen und koordinieren kann.

Auf Stufe 1A werden zunächst nur die Perspektiven unvermittelt nebeneinanderstellt (s.o.), jedem aber zugleich das Seine zugestanden. Auf Stufe 1B teilt man mit dem anderen oder wechselt sich bei bestimmten Handlungen ab (etwa beim Spiel mit einem Spielzeug oder wenn man, wie beim Diktatorspiel, einen gemeinsamen Betrag aufteilen soll). Hier greift „Inequity Aversion“ bzw. Reziprozität in ihrer einfachsten Form. Stufe

1C beschreibt hingegen das, was man allgemein als Altruismus bezeichnet und wofür Andreoni (vgl. 1990) das Konzept des „Warm Glow“ eingeführt hat, das zusätzlich unterstreicht, dass man Bedürftigen gerne hilft, wenn man besser ausgestattet ist und entsprechend helfen kann.

Eigeninteresse auf Stufe 2A geht darüber insofern hinaus, als es altruistische Interessen mit einschließt, aber zugleich der Tatsache Rechnung trägt, dass man nicht immer und nicht jedem helfen will und der Altruismus somit begrenzt ist. Nach diesem Prinzip soll jeder seine, wie auch immer gearteten, Interessen verfolgen.

Stufe	Moralische Stufen/Prinzipien
1A	Eigentum; Anerkennung des Willens und der Bedürfnisse anderer
1B	Mit anderen teilen bzw. sich abwechseln
1C	Bedarfsgerechte Verteilung (bei unterschiedlicher Bedürftigkeit)
2A	Interessenorientierung (wobei Interessen altruistisch sein können)
2B	Aushandlung/wechselseitige Versprechen
2C	Goldene Regel („Was du nicht willst, das man dir tu“, das füg’ auch keinem andern zu“)
...	...

*Abbildung 2: Moralkognitive Stufenprinzipien und soziale Präferenzen  
(Quelle: eigene Darstellung).*

Reziprozität im Sinne einer Aushandlung und wechselseitigen Koordination eigeninteressierter Akteure (Stufe 2B) tritt beispielsweise im mehrstufigen „Gift Exchange Game“ (vgl. Akerlof 1982) auf, oder aber auch im mehrstufigen Gefangenendilemma (vgl. Axelrod 1984/2005). Abgesehen vom wechselseitigen Desinteresse (wie bei typischen Käufen und Verkäufen), spielt hierbei eine Rolle, dass die individuellen Wert-schätzungen der gegebenen bzw. empfangenen Güter (also die damit verbundenen Interessen) interindividuell verschieden sein können; sie werden durch die Aushandlung und den Vertragsschluss jedoch kommensurabel gemacht.

Über diese Form der Reziprozität hinaus geht es schließlich, wenn man etwa als Arbeitgeber im Gift Exchange Game das gibt, was man für fair im Sinne der goldenen Regel hält, bzw. umgekehrt als Arbeitnehmer die Leistung erbringt, die man (evtl. angesichts der Lohnhöhe) für angemessen hält (Stufe 2C). Hier schließt man keinen Vertrag mit



dem jeweils anderen, sondern mit sich selbst aus einer übergreifenden Perspektive, von der aus man beide Interessensperspektiven zu integrieren versucht. Es handelt sich somit um eine übergreifende Fairness-Perspektive, wie sie z.B. in der „Fair wage-effort“-Hypothese (vgl. Akerlof/Yellen 1990) zum Ausdruck gebracht wird (vgl. hierzu auch Sobel 2005: 414–416).

Festzuhalten bleibt, dass solche Moralprinzipien die stabilen fundamentalen Präferenzen im Sinne von Dietrich und List<sup>10</sup> sind, und dass für sie aufgrund der hierarchischen Inklusion der jeweils niedrigeren in den jeweils höheren Stufen auch die relative Gewichtung der Moralprinzipien festgelegt ist.<sup>11</sup> Das muss allerdings nicht heißen, dass man stets nach dem höchsten moralischen Prinzip urteilt (wie Kohlberg unterstellt hat), sondern nur nach dem höchsten der situationsspezifisch relevanten Prinzipien. Ein Aspekt von Stabilität wäre dabei allerdings auch, dass in vergleichbaren Situationen die gleichen Prinzipien angesprochen werden müssten und daraus auch die gleichen Präferenzen (bzgl. der Alternativen) abgeleitet werden müssten (vgl. Dietrich/List 2013b: 622f.). Wie man die situationsspezifische Nutzung von Moralprinzipien modellieren kann, wird im folgenden Abschnitt beschrieben.

### 3.3 Situationsspezifität und Rationalität

Dass situationsspezifisch unterschiedliche Prinzipien genutzt werden, ist in der Forschung sehr gut belegt (vgl. Beck/Parche-Kawik 2004; Krebs/Denton 2005; Lindenberg 2008, Kaplan/Tivnan 2014). Früher ging man davon aus, dass es sich hierbei um sozialisationsbedingte Segmentierungen handelt, denen keine „moralische Vernunft“ zugrunde liegt (vgl. hierzu etwa Lempert 1988; Beck et al. 1996). Diese Sichtweise wurde jedoch vor allem im Lichte des Homann'schen Ansatzes einer Revision unterzogen und im Sinne einer situationsspezifischen Rationalität moralischer Differenzierungen neu interpretiert (vgl. Beck 2003; Minnameier 2004a; Minnameier/Beck 2015). Die obigen Ausführungen zu Framing-Effekten im Kontext der RCT unterstützen diese Interpretation.

Nach dieser Auffassung können und sollen Individuen entsprechend ihrer moralischen Entwicklung (im Sinne der verfügbaren stufenbezogenen Moralkonzepte) und konkreter verinnerlichter Werte (wie z.B. Zuverlässigkeit, Wahrhaftigkeit etc.), die heute unter dem Konzept des „moralischen Selbst“ (vgl. z.B. Johnston/Krettenauer 2011) oder der moralischen Identität (vgl. z.B. Shao et al. 2008; Bénabou/Tirole 2011) diskutiert werden, flexibel auf moralische Anforderungen reagieren (vgl. Abbildung 3; vgl. hierzu insbes. Minnameier 2012a; 2013b). Die Frage ist freilich, wie man sich die Prozesse einer solchen flexiblen Anpassung vorstellen muss und inwiefern diese als „rational“ aufzufassen sind.

---

<sup>10</sup> Dietrich und List sagen dies deutlich: „The stable feature characterizing an agent is not the agent's preference order over the alternatives in  $X$ , as in standard rational choice theory, but the agent's weighing relation over property combinations“ (2013b: 621f.). Man muss im vorliegenden Kontext nur „property combinations“ durch Moralprinzipien ersetzen.

<sup>11</sup> Es ist allerdings zu beachten, dass mit Moralprinzipien hier nicht einfache Imperative wie „Du sollst nicht stehlen!“ oder „Die Todesstrafe sollte verboten sein!“ gemeint sind, sondern Prinzipien, mit denen sich solche Forderungen begründen lassen.

Wie oben ausgeführt, verstehe ich Moralprinzipien als abstrakte Konzepte, die für bestimmte Handlungsentscheidungen (als Entscheidung zwischen Alternativen) relevant sein können. Und wenn sie das sind, lässt sich aus ihnen eine konkrete Präferenzordnung hinsichtlich der Handlungsalternativen ableiten. Insbesondere sind diejenigen Alternativen vorzuziehen, die die Kriterien des Prinzips erfüllen (bzw. sind diejenigen Alternativen auszuschließen, die es verletzen).

Welche Prinzipien relevant sind, hängt von der Entscheidungssituation (d.h. der moralischen Problemsituation) ab, in der sich der Akteur befindet. Zu den Situationsmerkmalen gehören nach Dietrich und List die konkreten Entscheidungsoptionen, aber auch relevante Framinginformationen, soziale Rollen, in denen sich der Akteur befindet, sowie seine konkreten Lebensumstände (vgl. 2013b: 622).<sup>12</sup> Ausgehend davon unterstellen sie einen funktionalen Zusammenhang, sodass aus einer bestimmten Situationskonstellation (also eines Situationstyps) bestimmte motivierende Gründe – also im vorliegenden Fall: Moralprinzipien – aktiviert werden (ebd.: 623).

Nehmen wir z.B. das klassische Diktatorspiel: Der Diktator ist in der Rolle entscheiden zu müssen, wie viel eines Geldbetrags er dem passiven Spieler zuweist. Die Optionen bestehen bei einer Ausstattung von 10€ und der Beschränkung auf ganzzahlige Eurobeträge genau in der Menge aller ganzzahligen Beträge zwischen und 0€ und 10€ für den Diktator und der entsprechenden komplementären Beträge für den passiven Spieler. Die eigenen Lebensumstände (soweit überhaupt relevant) sind dem Diktator bekannt, die seines Partners kennt er nicht.

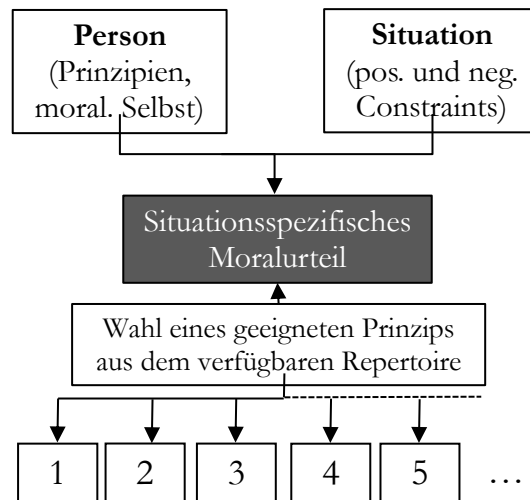


Abbildung 3: Situationsspezifische Nutzung moralischer Urteilsprinzipien  
(Quelle: eigene Darstellung)

<sup>12</sup> Dietrich und List fokussieren hier auf objektiv beobachtbare Situationsmerkmale, weil sie an dieser Stelle die Frage der Falsifizierbarkeit ihres Ansatzes diskutieren. Für die Modellierung der individuellen Verhaltensgenese käme es jedoch auf die *subjektive Situationskonstitution* an.

Ein moralisches Problem liegt vor, weil die Entscheidung notwendig den Partner affiziert. Und es ist, zumindest auf den ersten Blick, ein recht simples moralisches Problem. Ein hier relevantes Prinzip wäre das des Teilens, und so wird hälftiges Teilen auch üblicherweise als die „faire“ Auszahlung angesehen. Das entspricht moraltaxonomisch der Stufe 1B (gleichmäßiges Aufteilen, auf Basis des mitfühlenden Hineinversetzens in den Anderen, was einer reziproken Koordination beider Perspektiven entspricht).<sup>13</sup> Auf Stufe 1A käme Aufteilen nur in Frage, wenn von vornherein bestimmt würde, dass es sich um gemeinsames Eigentum handelt und dem Anderen insofern die 5€ „zustehen“. Stufe 1C hingegen ist insofern nicht relevant, als man ja gar nicht wissen kann, ob der Andere evtl. signifikant ärmer ist als man selbst oder umgekehrt.

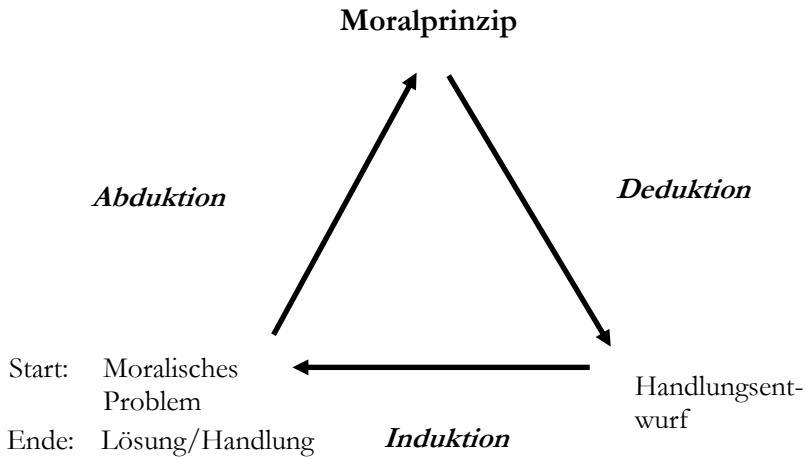
Wird die moralische Situation so konstituiert, und wird entsprechend das Prinzip der Stufe 1B aktiviert, so würde sich daraus eine Präferenz ableiten, die Hälfte abzugeben. Die Frage ist allerdings, ob sich eine „Präferenz“ sozusagen direkt *deduktiv* aus dem Prinzip und den situativen Prämissen ergibt. Dietrich und List präsentieren hier einen axiomatischen Ansatz, bei dem die situativen Voraussetzungen und sämtliche Randbedingungen definitorisch fixiert sind, und nur in diesem analytisch geschlossenen Rahmen determinieren Gründe und Randbedingungen die konkreten Präferenzen.<sup>14</sup> Man könnte beim Diktatorspiel aber z.B. überlegen, ob man das Geld nicht vielleicht dringend selbst benötigt, oder sich die Frage stellen, ob man wirklich moralisch verpflichtet ist, die Hälfte abzugeben, wo man doch sonst auch nicht alles mit jedem teilt. Damit wären zusätzliche Aspekte zu berücksichtigen, die bislang außer Acht gelassen wurden.

Es ist heute klar, dass weder lebensweltliche noch wissenschaftliche Reasoning-Prozesse rein deduktiv verlaufen, sondern sog. abduktive und induktive Inferenzen umfassen (vgl. Gabbay/Woods 2005; Magnani 2009; Minnameier 2004b; 2016). In Ergänzung und Erweiterung des Ansatzes von Dietrich und List schlage ich daher eine Modellierung auf Basis der Peirce'schen Inferenzen der Abduktion, Deduktion und Induktion vor (vgl. Abbildung 4). Sie sind nicht nur im Kontext der Wissensgenese, sondern auch im Kontext der Wissensanwendung und damit auch der Anwendung moralischer Prinzipien bedeutsam (vgl. Minnameier 2012a; 2013a; 2016).

---

<sup>13</sup> Reziprok im Übrigen auch deshalb, weil im Falle „verdienter Beträge“ nicht hälftig, sondern nach Leistung aufzuteilen wäre, ebenso wie nach dem Konzept der „Inequity Aversion“.

<sup>14</sup> Dietrich und List (2013a: 114f.) arbeiten im Übrigen auch heraus, dass die Gewichtungrelation selbst nicht vollständig und transitiv sein muss, was schon als solches die Möglichkeit einer strikt deduktiven Determination konkreter Präferenzen für bestimmte Alternativen limitiert.



*Abbildung 4: Moralisches Urteilen und Handeln im inferentiellen Zusammenhang  
(Quelle: eigene Darstellung)*

Kurz gesagt, lässt sich der inferentielle Zusammenhang wie folgt erläutern (für eine ausführliche Darstellung vgl. Minnameier 2013a; für den nicht-moralischen Zusammenhang vgl. Minnameier 2004b; 2010b; 2016): Die *Abduktion* ist der Prozess, der von einem wahrgenommenen Problem zu einer *möglichen* Lösung führt, hier von einem wahrgenommenen moralischen Problem im Sinne konkurrierender Ansprüche zu einem Prinzip, nach dem diese geregelt werden könnten. Die Abduktion kann damit zugleich als der Schritt angesehen werden, über den die situationsspezifische Aktivierung moralischer Urteilsprinzipien vermittelt ist. Abduktiv gültig ist ein Schluss, wenn er das anfängliche Problem zum Verschwinden bringt, sprich hier: das wahrgenommene moralische Problem prinzipiell zu lösen erlaubt. Aus dem so aktivierten Prinzip lässt sich deduktiv ableiten, was zu tun wäre, und dieser konkrete Handlungsentswurf wird im letzten (induktiven) Schritt evaluiert. Diese Evaluation ergibt im positiven Fall, dass die betreffende Handlung bzw. der Handlungstyp für alle Situationen des betreffenden Typs adäquat sind (induktive Generalisierung). Allerdings wird das so gewonnene Wissen mit jeder nachfolgenden Anwendung zugleich erneut geprüft, und das heißt: entweder bekräftigt oder aber auch problematisiert (wie oben am Beispiel des Diktatorspiels erläutert).

Die Bildung einer konkreten Präferenz im Sinne Dietrichs und Lists verlangt insofern über die deduktive Ableitung einer Handlung – als Konsequenz aus der Anwendung eines Moralprinzips auf die zuvor konstituierte Situation – hinaus daher streng genommen stets noch eine pragmatische Akzeptanz (induktiver Schluss), nämlich dass diese Konsequenz auch für diese konkrete Situation (die ja trotz aller Situationstypisierung immer auch einzigartig ist) gelten soll.

Die Darstellung in Abbildung 4 suggeriert, dass moralisches Urteilen und Handeln im inferentiellen Sinn nur für „System 2“ sensu Kahneman gilt. Es lässt sich jedoch auch

auf „System 1“ übertragen (vgl. hierzu Magnani 2009; Minnameier 2016). Es bedürfte freilich einer ausführlichen Erörterung um herauszuarbeiten, inwiefern intuitives Handeln ebenfalls auf inferentieller Basis zu erklären ist, die im vorliegenden Kontext allerdings nicht geleistet werden kann. Man denke in diesem Zusammenhang aber auch an Vernon Smiths Rationalitätskonzept, das sich nicht nur auf Kahnemans System 2 bezieht, sondern auch das Denken und Handeln im System 1 umfasst. Ganz in diesem Sinne unterscheidet Smith (vgl. 2008) zwischen einer konstruktivistischen und einer ökologischen Rationalität.

### 3.4 (Moralische) Rationalität und Irrationalität

Fassen wir die bisherigen Ergebnisse zusammen:

1. Nach der RBT werden aus Situationswahrnehmungen heraus bestimmte Prinzipien bzw. „Gründe“ aktiviert. Aus den Prinzipien und den spezifischen situativen Gegebenheiten werden schließlich konkrete Präferenzen abgeleitet.
2. Die Menge aller möglichen Prinzipien/Gründe und der zwischen ihnen bestehenden Gewichtungsrelation beschreibt die Struktur der *stabilen fundamentalen Präferenzen*. Für den Bereich der Moral wurde ergänzt, dass eine Hierarchie sukzessive aufeinander aufbauender und einander integrierender Prinzipien denkbar ist, so dass das jeweils höchste der situationsspezifisch relevanten Prinzipien am Ende die konkrete Präferenz bestimmt.<sup>15</sup>
3. Die *Restriktionen* sind im Wesentlichen durch die Situationskonstitution gegeben, können jedoch im Rahmen der induktiven Evaluation durch zusätzliche Aspekte ergänzt werden.
4. Die am Ende abgeleiteten *konkreten Präferenzen* sind die resultierenden situationsspezifischen Handlungsentscheidungen. Sie ergeben sich aus den fundamentalen Präferenzen im Lichte der insgesamt wahrgenommenen Restriktionen.

Nun ist allerdings darauf hinzuweisen, dass *Handlungsentscheidungen* nicht das gleiche sind wie *Handlungen*. Deshalb sind auch die situationsspezifisch gebildeten Präferenzen nicht gleichbedeutend mit „*Choices*“. Darauf weisen Dietrich und List explizit hin:

„(T)he way decision theory formally relates preferences to choices carries over to reason-based preferences. Any preference order, including a reason-based

<sup>15</sup> In diesem Zusammenhang wurde auch angemerkt, dass Stabilität bzw. Fundamentalität von Prinzipien nicht absolut zu verstehen ist, sondern relativ zu den Entscheidungsobjekten, über die eine konkrete Präferenz zu bilden ist. Das schließt aber eben nicht aus, dass Prinzipien ihrerseits zum Gegenstand der Reflexion gemacht werden. Das gilt speziell dann, wenn sich mithilfe eines Prinzips keine eindeutige konkrete Präferenz bestimmen lässt. *Beispiel*: Betrachtet man die Situation des Gefangenendilemmas aus der integrativen Perspektive der Stufe 1C, so fühlt man einerseits mit dem Anderen mit, zugleich aber ebenso mit sich selbst. Es lässt sich aber keine entsprechende integrative Entscheidung ableiten. Diese „Situation“ kann entsprechend zum Prinzip der Stufe 2A führen, wonach jeder Akteur nach seinem subjektiven Interesse entscheiden muss (wobei im Übrigen zu bedenken ist, dass das Interesse ausgeprägter Altruisten darin bestehen kann, in jedem Fall zu kooperieren).

one, induces a corresponding choice function. While a preference order represents certain *intentional attitudes* towards a given set of alternatives, a choice function encodes certain *choice dispositions* in relation to these alternatives. It specifies which alternative, or alternatives, would (or on a normative interpretation, should) be chosen from any available subset of the alternatives.” (Dietrich/List, 2013a: 125)

Rational ist demnach ein Handeln, das den gebildeten Präferenzen folgt. Aber das muss ja nicht so geschehen. Es handelt sich vielmehr um einen separaten Akt der Umsetzung der gebildeten Handlungsintention, und eine solche Umsetzung kann stets scheitern. Für ein solches Scheitern können Unfähigkeit oder Unaufmerksamkeit die Ursache sein, im moralischen Zusammenhang kommt außerdem Willensschwäche bzw. mangelnde Ich-Kontrolle hinzu. Solche Divergenzen zwischen Handlungsentscheidungen (*Preferences*) und Handlungen (*Choices*) konstituieren *Irrationalität* und führen entsprechend zu schlechtem Gewissen, Bedauern, Selbstvorwürfen und dergleichen mehr.

Ein zusätzlicher Aspekt von Irrationalität, der allerdings einer eigenständigen detaillierten Erörterung bedürfte, kommt ins Spiel, wenn man berücksichtigt, dass die in Abschnitt 3.3 diskutierten Inferenzen fehlerhaft sein können (vgl. Minnameier 2012b). Hier würden konkrete Präferenzen gebildet, die nicht den zugrunde liegenden moralischen Prinzipien und sonstigen relevanten Prämissen entsprechen.

So erlaubt die RBT einerseits eine lückenlose Erklärung menschlichen Verhaltens (wie sie z. B. Gary Becker beansprucht hat), gestattet aber zugleich, sowohl analytisch als auch empirisch zwischen rationalem und irrationalem Handeln zu unterscheiden.

#### 4. Schlussbemerkung

Bruni und Sugden (2007) zeigen einen zentralen Konflikt zwischen der RCT-basierten Spieltheorie und der Verhaltensökonomik auf:

„Game theory, interpreted as modelling the strategic interactions of ideally rational agents, was seen as the logical completion of rational-choice theory; there was a widespread hope that, by providing a universal theory of choice, game theory would unify the social sciences. Against this background, behavioural economics stands out as a counter-revolution.” (ebd.: 146f.)

Die fundamentale Frage, um die es ihrer Ansicht nach geht, lautet demzufolge: „(S)hould economic theory be based on assumptions about human psychology or on assumptions about rational choice?“ (ebd.: 147). Der vorliegende Beitrag zeigt m.E., dass diese Dichotomie im Lichte der RBT und der darauf aufbauenden Überlegungen überwunden werden kann, und dass eine solche Überwindung für die Wirtschaftsethik von zentraler systematischer Bedeutung ist.

#### Literaturverzeichnis

- Akerlof, G. A. (1982): Labor Contracts as a Partial Gift Exchange, in: The Quarterly Journal of Economics, Vol. 97/No. 4, 543–569.
- Akerlof, G. A./Yellen, J. L. (1990): The fair Wage-Effort Hypothesis and Unemployment, in: The Quarterly Journal of Economics, Vol. 105/No. 2, 255–283.

- Andreoni, J.* (1990): Impure Altruism and Donations to Public Goods: A Theory of Warm-Glow Giving, in: *Economic Journal*, Vol. 100/No. 401, 464–477.
- Andreoni J./Bernheim D. B.* (2009): Social Image and the 50–50 Norm: A Theoretical and Experimental Analysis of Audience Effects, in: *Econometrica*, Vol. 77/No. 5, 1607–1636.
- Arrow, K. J.* (1972): Gifts and Exchanges, in: *Philosophy and Public Affairs*, Vol. 1/No. 4, 343–362.
- Axelrod, R.* (1984/2005): *Die Evolution der Kooperation*, München: Oldenbourg.
- Bartling, B./Fischbacher, U.* (2012): Shifting the Blame: On Delegation and Responsibility, in: *Review of Economic Studies*, Vol. 79/No. 1, 67–87.
- Beck, K./Parche-Kawik, K.* (2004): Das Mäntelchen im Wind? Zur Domänenspezifität moralischen Urteilens, in: *Zeitschrift für Pädagogik*, Jg. 50/Heft 2, 244–265.
- Beck, K.* (2003): Ethischer Universalismus als moralische Verunsicherung? Zur Diskussion um die Grundlegung der Moralerziehung, in: *Zeitschrift für Berufs- und Wirtschaftspädagogik*, Jg. 99/Heft 2, 274–298.
- Beck, K./Dransfeld, A./Minnameier, G./Wuttke, E.* (2002): Autonomy in Heterogeneity? Development of Moral Judgement Behaviour during Business Education, in: Beck, K. (Ed.): *Teaching-Learning Processes in Vocational Education. Foundations of Modern Training Programmes*, Frankfurt: Peter Lang, 87–119.
- Beck, K./Brütting, B./Liidecke, S./Minnameier, G./Schirmer, U./Schmid, S. N.* (1996): Zur Entwicklung moralischer Urteilskompetenz in der kaufmännischen Erstausbildung – Empirische Befunde und praktische Probleme, in Beck, K./Heid, H. (Hrsg.): *Lehr-Lern-Prozesse in der kaufmännischen Erstausbildung – Wissenserwerb, Motivierungsgeschehen und Handlungskompetenzen*, Stuttgart: Steiner, 187–206.
- Becker, G. S.* (1976): Altruism, Egoism, and genetic Fitness: Economics and Sociobiology, in: *Journal of Economic Literature*, Vol.14/No.3, 817–26.
- Becker, G. S.* (1993): Nobel Lecture: The Economic Way of Looking at Behavior, in: *Journal of Political Economy*, Vol. 101/No. 3, 385–409.
- Becker, G. S.* (1996): *Accounting for Tastes*, Cambridge (MA): Harvard University Press.
- Becker, G. S./Stigler, G. J.* (1977): De Gustibus non est Disputandum, in: *American Economic Review*, Vol. 67/No. 2, 76–90.
- Benabou, R./Tirole, J.* (2011): Identity, Moral, and Taboos: Beliefs as Assets, in: *Quarterly Journal of Economics*, Vol. 126/No. 2, 805–855.
- Binmore, K./Shaked, A.* (2010): Experimental Economics: Where next?, in: *Journal of Economic Behavior and Organization*, Vol. 73/No. 1, 87–100.
- Binmore, K.* (2009): *Rational Decisions*, Princeton (NJ): Princeton University Press.
- Birnbacher, D.* (2007): *Analytische Einführung in die Ethik*, 2. Aufl., Berlin: de Gruyter.
- Bolton, G. E./Ockenfels, A.* (2000): ERC: A Theory of Equity, Reciprocity and Competition, in: *American Economic Review*, Vol. 90/No. 1, 166–193.
- Bruni, L./Sugden, R.* (2007): The Road Not Taken: How Psychology Was Removed from Economics and How It Might Be Brought Back, in: *The Economic Journal*, Vol. 117/No. 516, 146–173.
- Cohn, A./Febr, E./Maréchal, A.* (2014): Business Culture and Dishonesty in the Banking Industry, in: *Nature*, Vol. 516/No. 4, 86–89.

- Colby, A./Kohlberg, L. (1987): The Measurement of Moral Judgement – Theoretical Foundations and Research Validation, Vol. I., Cambridge (UK): Cambridge University Press.
- Dana, J./Weber, R. A./Kuang, J. X. (2007): Exploiting Moral Wiggle Room: Experiments Demonstrating an Illusory Preference for Fairness, in: *Economic Theory*, Vol. 33/No. 1, 67–80.
- Dietrich, F./List, D. (2013a): A Reason-based Theory of Rational Choice, in: *Noûs*, Vol. 47/No. 1, 104–134.
- Dietrich, F./List, D. (2013b): Where Do Preferences Come From?, in: *International Journal of Game Theory*, Vol. 42/No. 3, 613–637.
- Dreber, A./Ellingsen, T./Johannesson, M./Rand, D. G. (2013): Do People Care About Social Context? Framing Effects in Dictator Games, in: *Experimental Economics*, Vol. 16/No. 3, 349–371.
- Dufvenberg, M./Kirchsteiger, G. (2004): A Theory of Sequential Reciprocity, in: *Games and Economic Behavior*, Vol. 47/No. 2, 268–298.
- Ellingsen, T./Johannesson, M./Mollerstrom, J./Munkhammar, S. (2012): Social Framing Effects: Preferences or Beliefs?, in: *Games and Economic Behavior*, Vol. 76/No. 1, 117–130.
- Engelmann, D./Strobel, M. (2004): Inequality Aversion, Efficiency, and Maximin Preferences in Simple Distribution Experiments, in: *American Economic Review*. Vol. 94/No. 4, 857–869.
- Falk, A./Fischbacher, U. (2006): A Theory of Reciprocity, in: *Games and Economic Behavior*, Vol. 54/No. 2, 293–315.
- Falk, A./Febr, E./Fischbacher, U. (2008): Testing Theories of Fairness – Intentions matter, in: *Games and Economic Behavior*, Vol. 62/No. 1, 287–303.
- Febr, E./Schmidt, K. M. (1999): A Theory of Fairness, Competition, and Cooperation, in: *Quarterly Journal of Economics*, Vol. 114/No. 3, 817–868.
- Febr, E./Schmidt, K. M. (2006): The Economics of Fairness, Reciprocity and Altruism: Experimental Evidence and New Theories, in Kolm, S./Ythier, J. (Eds): *Handbook on the Economics of Giving, Reciprocity, and Altruism*, Vol. 1, Amsterdam: Elsevier, 615–669.
- Forsythe, R./Horowitz, J./Savin, N./Sefton, M. (1994): Fairness in Simple Bargaining Experiments, in: *Games and Economic Behavior*, Vol. 6/No. 3, 347–369.
- Gabbay, D. M./Woods, J. (2005): *A Practical Logic of Cognitive Systems*, Vol. 2: The Reach of Abduction – Insight and Trial, Amsterdam: Elsevier.
- Gächter, S. (2013): Rationality, Social Preferences, and Strategic Decision-Making, in: Wittek, R./Snijders, T./Nee, V. (Eds.): *The Handbook of Rational Choice Social Research*, Palo Alto (CA): Stanford University Press, 33–71.
- Hamman, J. R./Loewenstein, G./Weber, R. A. (2010): Self-Interest through Delegation: An Additional Rationale for the Principal-Agent Relationship, in: *American Economic Review*, Vol. 100/No. 4, 1826–1846.
- Henrich, J./Boyd, R./Bowles, S./Camerer, C./Febr, E./Gintis, H. (Eds.) (2004): *Foundations of Human Sociality: Economics Experiments and Ethnographic Evidence from Fifteen Small-Scale Societies*, New York: Oxford University Press.
- Hoffman, E./McCabe, K./Smith, V. (1996): Social Distance and Other-regarding Behavior, in: *American Economic Review*, Vol. 86/No. 3, 653–660.
- Hollis, M./Sugden, R. (1993). Rationality in Action, in: *Mind*, Vol. 102/No. 405, 1–35.



- Homann, K. (1994): Ethik und Ökonomik: Zur Theoriestrategie der Wirtschaftsethik, in: Homann, K. (Hrsg.): Wirtschaftsethische Perspektiven I – Theorie, Ordnungsfragen, Internationale Institutionen, Berlin: Duncker/Humboldt, 9–30.
- Homann, K. (2001): Ökonomik: Fortsetzung der Ethik mit anderen Mitteln, in: Siebeck, G. (Hrsg.): *Artibus Ingeniis – Beiträge zu Theologie, Philosophie, Rechtswissenschaft und Wirtschaftswissenschaft*, Tübingen: Mohr Siebeck, 85–110.
- Homann, K. (2004): Braucht die Wirtschaftsethik eine „moralische Motivation“?, in: Arnold, V. (Hrsg.): *Wirtschaftsethische Perspektiven VII: Methodische Grundsatzfragen, Unternehmensethik, Verteilungsfragen, Gentechnik und Fragen der medizinischen Ethik*, Berlin: Duncker/Humboldt, 33–60.
- Homann, K. (2014): *Sollen und Können: Grenzen und Bedingungen der Individualmoral*, Wien: Ibra.
- Homann, K./Suchanek, A. (2005): *Ökonomik: Eine Einführung*, 2. Aufl., Tübingen: Mohr Siebeck.
- Johnston, M./Krettenauer, T. (2011): Moral Self and Moral Emotion Expectancies as Predictors of Anti- and Prosocial Behavior in Adolescence: A Case for Mediation?, in: *European Journal of Developmental Psychology*, Vol. 8/No. 2, 228–243.
- Kahneman, D. (2012): *Thinking, Fast and Slow*, London: Penguin Books.
- Kaplan, U./Tivnan, T. (2014): Moral Motivational Pluralism: Moral Judgment as a Function of the Dynamic Assembly of Multiple Developmental Structures, in: *Journal of Adult Development*, Vol. 21/No 4, 193–206.
- Kirchgässner, G. (2013): Das schwache Rationalitätsprinzip als Grundlage einer verstehenden Sozialwissenschaft, in: *Zeitschrift für Wirtschafts- und Unternehmensethik*, Jg. 14/Heft 2, 259–281.
- Kohlberg, L./Levine, C./Hewer, A. (1984): The Current Formulation of the Theory, in: Kohlberg, L.: *Essays on Moral Development*, Vol. II, *The psychology of Moral Development. The Nature and Validity of Moral Stages*, San Francisco (CA): Harper/Row, 207–319.
- Kohlberg, L. (1997): *Die Psychologie der Moralentwicklung* (hrsg. v. Althof, W.), 2. Aufl., Frankfurt am Main: Suhrkamp.
- Krebs, D. L./Denton, K. (2005): Toward a More Pragmatic Approach to Morality: A critical Evaluation of Kohlberg's Model, in: *Psychological Review*, Vol. 113/No. 3, 629–649.
- Ledyard, J. (1995): Public Goods: A Survey of Experimental Research, in: Kagel, J./Roth, A. (Eds.): *Handbook of Experimental Economics*, Princeton: Princeton University Press, 253–279.
- Lempert, W. (1988): *Moralisches Denken – Seine Entwicklung jenseits der Kindheit und seine Beeinflussbarkeit in der Sekundarstufe II*, Essen: Neue Deutsche Schule.
- Levitt, S./List, J. (2007): What do Laboratory Experiments Measuring Social Preferences Reveal about the Real World?, in: *Journal of Economic Perspectives*, Vol. 21/No. 2, 153–174.
- Lieberman, V./Samuels, S. M./Ross, L. (2004): The Name of the Game: Predictive Power of Reputations Versus Situational Labels in Determining Prisoner's Dilemma Game Moves, in: *Personality and Social Psychology Bulletin*, Vol. 30/No. 9, 1175–1185.
- Lindenberg, S. (2008): Social Norms: What Happens When They Become More Abstract?, in: Diekmann/A., Eichner, K./Schmidt, P./Voss, T. (Hrsg.): *Rational choice: Theoretische Analysen und empirische Resultate*, Wiesbaden: VS Verlag, 63–81.

- List, J. A. (2007): On the Interpretation of Giving in Dictator Games, in: Journal of Political Economy, Vol.115/No. 3, 482–494.
- Magnani, L. (2009): Abductive Cognition: The Epistemological and Eco-Cognitive Dimensions of Hypothetical Reasoning, Berlin: Springer.
- Minnameier, G. (2000): Strukturgenese moralischen Denkens: Eine Rekonstruktion der Piaget-schen Entwicklungslogik und ihre moraltheoretischen Folgen, Münster: Waxmann.
- Minnameier, G. (2004a): Ethics and Economics, Friends or Foes? An Educational Debate, in: Journal of Moral Education, Vol. 33/No. 3, 359–369.
- Minnameier, G. (2004b): Peirce-Suit of Truth: Why Inference to the Best Explanation and Abduction Ought Not to be Confused, in: Erkenntnis, Vol. 60/No. 1, 75–105.
- Minnameier, G. (2005): Developmental Progress in Ancient Greek Ethics, in: European Journal of Developmental Psychology, Vol. 2/No. 1, 71–99.
- Minnameier, G. (2010a): Entwicklung moralischen Denkens aus einer neo-kohlbergschen Perspektive, in: Latzko, B./Malti, T. (Hrsg.): Moralentwicklung und –erziehung in Kindheit und Adoleszenz, Göttingen: Hogrefe, 47–67.
- Minnameier, G. (2010b): The Logicity of Abduction, Deduction, and Induction, in: Bergman, M./Paavola, S./Pietarinen, A. V./Rydenfelt, H. (Eds.): Ideas in Action: Procedures of the Applying Peirce Conference, Helsinki: Nordic Pragmatism Network, 216–228.
- Minnameier, G. (2012a): A Cognitive Approach to the ‘Happy Victimiser’, in: Journal of Moral Education, Vol. 41/No. 4, 491–508.
- Minnameier, G. (2012b): What’s Wrong With It? – Kinds and Inferential Mechanics of Reasoning Errors, in: Seifried, J./Wuttke, E. (Eds.): Learning from Errors; Opladen: Verlag Barbara Budrich, 13–29.
- Minnameier, G. (2013a): Deontic and Responsibility Judgments: An Inferential Analysis, in: Oser, F. K./Heinrichs, K./Lovat, T. (Eds.): Handbook of Moral Motivation: Theories, Models, Applications, Rotterdam: Sense, 69–82.
- Minnameier, G. (2013b): Der homo oeconomicus als „happy victimizer“: Strukturen und Prozesse moralischen Funktionierens im verhaltensökonomischen Kontext, in: Zeitschrift für Wirtschafts- und Unternehmensethik, Jg. 14/Heft 2, 119–135.
- Minnameier, G. (2016): Forms of Abduction and an Inferential Taxonomy, in: Magnani, L./Ber-tolotti, T. (Eds.): Springer Handbook of Model-based Science, Berlin: Springer, im Erscheinen.
- Minnameier G./Beck, K. (2015): Selbstorganisation und Situationsspezifität im Bereich der (Berufs-)Moral, in: Rausch, A./Warwas, J./Seifried, J./Wuttke, E. (Hrsg.), Konzepte und Ergebnisse ausgewählter Forschungsfelder der Berufs- und Wirtschaftspädagogik – Festschrift für Detlef Sembill, Baltmannsweiler: Schneider Verlag Hohengehren, 175–192.
- Minnameier, G./Schmidt, S. (2013): Situational Moral Adjustment and the Happy Victimizer, in: European Journal of Developmental Psychology, Vol. 10/No. 2, 253–268.
- Neuweg, G. H. (2015): Das Schweigen der Könnner: Gesammelte Aufsätze zum impliziten Wissen, Münster: Waxmann.
- Niekamp, M. (2014): Zum systematischen Verhältnis von Ordnungsethik und Verhaltensökonomik, in: Zeitschrift für Wirtschafts- und Unternehmensethik, Jg. 15/Heft 3, 335–350.

- Piaget, J./Garvia, R.* (1989): Psychogenesis and the History of Science, New York: Columbia University Press.
- Pies, I.* (Hrsg.) (2009): Moral als Heuristik: Ordonomische Schriften zur Wirtschaftsethik, Berlin: wvb.
- Pies, I./Hielscher, S.* (2014): Verhaltensökonomik versus Ordnungsethik? – Zum moralischen Stellenwert von Dispositionen und Institutionen, in: Zeitschrift für Wirtschafts- und Unternehmensethik, Vol. 15/No. 3, 398–420.
- Popper, K.* (1967/1997): Das Rationalitätsprinzip, in: Popper, K.: Lesebuch – Ausgewählte Texte zu Erkenntnistheorie, Philosophie der Naturwissenschaften, Metaphysik, Sozialphilosophie (hrsg. v. Miller, D.), Tübingen: Mohr Siebeck, 350–359.
- Rabin, M.* (1993): Incorporating Fairness into Game Theory and Economics, in: American Economic Review, Vol. 83/No. 9, 1281–1302
- Samuelson, P. A.* (1938): A Note on the Pure Theory of Consumer's Behaviour, in: *Economica*, Vol. 5/No. 17, 61–71.
- Samuelson, P. A.* (1948): Consumption Theory in Terms of Revealed Preference, in: *Economica*, Vol. 15/No. 60, 243–253.
- Savage, L.* (1954): The Foundations of Statistics, New York: Wiley.
- Sen, A. K.* (1977): Rational Fools: A Critique of the Behavioral Foundations of Economic Theory, in: *Philosophy and Public Affairs*, Vol. 6/No. 4, 317–344.
- Shao, R./Aquino, K./Freeman, D.* (2008): Beyond Moral Reasoning: A Review of Moral Identity Research and its Implications for Business Ethics, in: *Business Ethics Quarterly*, Vol. 18/No. 4, 513–540.
- Smith, V. L.* (2008): Rationality in Economics: Constructivist and Ecological Forms, Cambridge: Cambridge University Press.
- Sobel, J.* (2005): Interdependent Preferences and Reciprocity, in: *Journal of Economic Literature*, Vol. 43/No. 2, 392–436.
- Tversky, A./Kahneman, D.* (1986): Rational Choice and the Framing of Decisions, in: *Journal of Business*, Vol. 59/No. 4, 251–278.
- Vanberg, V. J.* (2008): Rationalitätsprinzip und Rationalitätshypothesen: Zum methodologischen Status der Theorie rationalen Handelns, in: Vanberg, V. J.: Wettbewerb und Regelordnung (hrsg. v. N. Goldschmidt, N./Wohlgemuth, M.), Tübingen: Mohr Siebeck, 215–230.