

# Willensschwäche aus handlungstheoretischer und moralphilosophischer Perspektive

---

Dissertation

zur Erlangung des Doktorgrades der Philosophie

an der Karl-Franzens-Universität Graz

eingereicht von

Richard Paul Hofmann

am Institut für Philosophie

Erstbegutachterin: Ao. Univ.-Prof. Dr. Sonja Rinofner-Kreidl

Zweitbegutachter: Univ.-Prof. Dr. Lukas Meyer

2013

# Zusammenfassung

---

Diese Arbeit ist eine systematische Untersuchung zu Phänomen und Begriff der Willensschwäche aus einer handlungstheoretischen und moralphilosophischen Perspektive. Ziel der Untersuchung ist ein besseres Verständnis von Willensschwäche hinsichtlich grundlegender Theorien des Handelns und Urteilens von Personen, hauptsächlich im Rahmen der gegenwärtig geführten philosophischen Debatte. In Kapitel I werden verschiedene Weisen der Begriffsbestimmung von Willensschwäche diskutiert und zwischen zwei wesentlichen Formen von Willensschwäche unterschieden: dem Handeln wider das eigene, bessere Urteil und der vorschnellen Revision von Handlungsintentionen. In Kapitel II wird in die zentralen philosophischen Probleme mit Willensschwäche anhand zweier historischer Diskussionsstränge, den Beiträgen von Aristoteles und Donald Davidson, eingeführt. In den Kapiteln III bis VI wird das Phänomen der Willensschwäche schließlich in seinem Bezug zu vier philosophischen Problemfeldern diskutiert: dem Verhältnis von Evaluation und Motivation, der Intentionalität von Handlungen, der Verantwortung für willensschwache Handlungen sowie der Normativität von Willensschwäche. Kapitel VII führt die wesentlichen Ergebnisse der Untersuchung in einer Deutung von Willensschwäche als Kehrseite unserer Fähigkeit zur reflexiven Selbstbestimmung zusammen.

## Abstract

---

This work provides a systematic enquiry of the phenomenon and concept of weakness of will within the perspective of action theory and moral philosophy. Aim of the enquiry is to gain a better understanding of weakness of will concerning fundamental theories of action and judgment of persons, mainly within the context of recent philosophical debates. Chapter I discusses different ways of conceptualizing weakness of will and distinguishes between two main forms of weakness of will: action against better judgment and the premature revision of an intention to act. Chapter II introduces the main philosophical problems surrounding weakness of will with the help of two historical strains of discussion, the contributions of Aristotle and Donald Davidson. Chapters III to VI discuss weakness of will within four central fields of philosophical problems: the relation between evaluation and motivation, the intentionality of actions, the responsibility for weak-willed actions and the normativity of weakness of will. Chapter VII brings together the different results in an analysis of weakness of will as the counterpart of our ability for reflexive self-determination.

# Vorwort

---

Ohne die Hilfe und Unterstützung von vielen Seiten wäre diese Arbeit nicht möglich gewesen. Sonja Rinofner-Kreidl hat mich zuerst für das Thema Willensschwäche begeistert und mich als Betreuerin der Arbeit immer wieder zur kritischen Überprüfung meiner Ideen angehalten und ermuntert. Lukas Meyer hat mir nicht nur die Chance gegeben, mein Talent unter Beweis zu stellen, sondern auch meinem philosophischen Denken neuen Schwung verliehen. Meinen Kolleginnen und Kollegen des Instituts für Philosophie und des Arbeitsbereiches für Praktische Philosophie an der Karl-Franzens-Universität Graz, insbesondere Alexa Zellentin, Pranay Sanklecha, Philippe Streit, Harald Stelzer, Claudia Reitingner, Amelie Stuart und Nora Kreft, danke ich für viele gemeinsame konstruktive Gespräche und kreative Zerstreuung auch außerhalb der Universität. Hilfreich waren auch die kritischen Fragen und Einwände zu Teilen meiner Arbeit seitens der Teilnehmerinnen und Teilnehmer der Seminare von Udo Thiel, Sonja Rinofner-Kreidl und Lukas Meyer sowie des Doktoratskolloquiums. Zudem bedanke ich mich bei den Kolleginnen und Kollegen aus Aachen und Wien, allen voran Wilfried Hinsch und Herlinde Pauer-Studer, für ihren Besuch in Graz und die gemeinsame Diskussion. Meinen Kolleginnen und Kollegen am Fachbereich für Sozial- und Wirtschaftswissenschaften (Rechts- und Sozialphilosophie) der Universität Salzburg, insbesondere Stephan Kirste, Silvia Augeneder und Günther Kreuzbauer, danke ich für die herzliche Aufnahme und wertvolle Diskussionen. Profitieren konnte ich auch von den Nachfragen zu meinen Vorträgen auf dem DissertantInnensymposium der Österreichischen Gesellschaft für Philosophie in Salzburg 2010 sowie auf der philosophischen Konferenz in Bled 2012. Ein großes Dankeschön geht auch an Jörn Müller, nicht nur für sein wunderbares Buch über Willensschwäche, sondern dass er sich auch die Zeit genommen hat, einen sommerlichen Vormittag in Würzburg Teile meiner Arbeit zu diskutieren. Die Dissertation wurde durch ein Forschungsstipendium und ein Doktoratsstipendium der Geisteswissenschaftlichen Fakultät der Karl-Franzens-Universität Graz unterstützt. Besonderer Dank gilt meinem Freund Dániel Bíró für die vielen gemeinschaftlichen Gespräche und Diskussionen sowie die Mühe des Korrekturlesens. Meinen Freunden und meiner Familie, Hannes, Reingard, Raimund und Reinhild, sowie Lorena Antón Fernandez danke ich für ihre Liebe und Unterstützung.

# Inhalt

---

<b>Einleitung .....</b>	<b>1</b>
<b>I Phänomen und Begriff der Willensschwäche .....</b>	<b>10</b>
I.1 Willensschwäche im Alltag, in der Kunst und als Gegenstand der Forschung .....	10
I.2 Formen und Typologien von Willensschwäche .....	12
I.2.1 Klassifikation von Willensschwäche nach Amélie Rorty .....	14
I.2.2 Ein interdisziplinäres Rahmenmodell für Willensschwäche .....	17
I.3 Akrasia oder Willensschwäche? .....	22
I.4 Willensschwache Überzeugungen .....	24
I.5 Der Begriff der Willensschwäche .....	25
I.5.1 Definition .....	25
I.5.2 Dimensionen von Willensschwäche .....	28
I.5.3 Abgrenzungen .....	30
Zusammenfassung .....	33
<b>II Willensschwäche als philosophisches Problem .....</b>	<b>35</b>
II.1 Akrasia bei Aristoteles .....	36
II.1.1 Die Diskussion der Akrasia innerhalb der NE .....	36
II.1.2 Aufbau von NE VII.1–11 .....	38
II.1.3 Interpretatorische Eckpunkte .....	41
II.1.4 Das Wissen des Akratikers: Interpretationen von NE VII.5 .....	44
II.2 Willensschwäche bei Donald Davidson .....	57
II.2.1 Kausale Handlungstheorie .....	58
II.2.2 Wie ist Willensschwäche möglich? .....	62
II.2.3 Paradoxien der Irrationalität .....	66
II.2.4 Probleme & Kritik I: Davidsons kausale Handlungstheorie .....	67
II.2.5 Probleme & Kritik II: Davidsons Konzeption von Willensschwäche .....	77
Zusammenfassung .....	83
<b>III Problemfeld Evaluation und Motivation .....</b>	<b>86</b>
III.1 Willensschwäche als Selbsttäuschung .....	86
III.1.1 Ursula Wolf: Zum Problem der Willensschwäche .....	86
III.1.2 Julius Schälike: Willensschwäche als Selbsttäuschung .....	92
III.2 Willensschwäche als fehlerhafte Diskontierung .....	99
III.2.1 Protagoras: Willensschwäche als fehlende Messkunst .....	100
III.2.2 George Ainslie: Willensschwäche als hyperbolische Diskontierung .....	103
III.3 Semantische Theorien und Sprechakttheorien .....	118
III.3.1 Richard Hare: Präskriptivismus und Willensschwäche .....	118
III.3.2 Robert Dunns Sprechakttheoretische Kritik .....	123

III.4 Willensschwäche, Inkommensurabilität und Inkomparabilität .....	128
III.4.1 David Wiggins: Inkommensurable Werte und Willensschwäche .....	128
III.4.2 Michael Stocker: Willensschwäche trotz Kommensurabilität .....	134
III.5 Motivationaler Internalismus/Externalismus und Willensschwäche .....	137
III.5.1 Motivationaler Internalismus .....	138
III.5.2 Urteils-Internalismus, Proportionalismus und Willensschwäche .....	148
III.5.3 Sergio Tenenbaum: Appearances of the good .....	152
III.5.4 Urteils-Externalismus und Willensschwäche .....	155
Zusammenfassung .....	160
<b>IV Problemfeld Intentionalität .....</b>	<b>165</b>
IV.1 Handlungstheoretische Grundlagen .....	166
IV.2 Handlungsleitende Intention und Willensschwäche .....	170
IV.2.1 Backward connection .....	170
IV.2.2 Forward connection .....	174
IV.2.3 Gründe-Theorie intentionalen Handelns .....	184
IV.3 Was sind praktische Gründe? .....	187
IV.3.1 Motivationaler Internalismus und Externalismus von praktischen Gründen .....	187
IV.3.2 Psychologische und anti-psychologische Theorien praktischer Gründe .....	188
IV.3.3 Wunsch- und wertebasierte Theorien praktischer Gründe .....	193
IV.4 Sind willensschwache Handlungen absichtlich? .....	195
IV.5 Willensschwäche und Pläne .....	206
IV.6 Exekutive Willensschwäche .....	210
Zusammenfassung .....	212
<b>V Problemfeld Verantwortung .....</b>	<b>214</b>
V.1 Begriff der Verantwortung .....	214
V.2 Kriterien für die Zuschreibung von Verantwortung .....	216
V.3 Willensschwäche und Zwang .....	221
Zusammenfassung .....	227
<b>VI Problemfeld Normativität .....</b>	<b>228</b>
VI.1 Willensschwäche als Form praktischer Irrationalität .....	228
VI.1.1 Interne Irrationalität von Willensschwäche .....	230
VI.1.2 Externe Irrationalität von Willensschwäche .....	236
VI.2 Willensschwäche als relatives Laster .....	239
VI.3 Willensschwäche als fehlendes Selbstvertrauen .....	243
VI.4 Willensschwäche und Selbstachtung .....	247
VI.5 Ist globale Willensschwäche möglich? .....	249
Zusammenfassung .....	251
<b>VII Willensschwäche als Kehrseite reflexiver Selbstbestimmung .....</b>	<b>253</b>
<b>Bibliographie .....</b>	<b>257</b>

*Und ich erkenne das Schreckliche, das ich zu tun gedenke,  
Doch mein Zorn ist stärker als meine Überlegungen,  
der schuld ist an den größten Übeln für die Sterblichen  
(Medea, Euripides)*

*Wohin mit den Millionen von Tatsachen, die da bezeugen,  
dass Menschen vorsätzlich, das heißt bei voller Einsicht in ihren wirklichen Vorteil,  
diesen dennoch hintansetzten und einen anderen Weg einschlugen...  
(Aufzeichnungen aus dem Kellerloch, Dostojewski)*

## Einleitung

In ihrer Einführung zum Schwerpunktthema Willensschwäche einer Ausgabe der *Deutschen Zeitschrift für Philosophie* (2009) benennen Juliane Rebentisch und Dirk Setton das generelle Unbehagen, das die philosophische Beschäftigung mit Willensschwäche begleitet:

„Jeder kennt es, und doch ist es rätselhaft; auf den ersten Blick erscheint es wie ein übersichtliches Spezialproblem aus der praktischen Philosophie, auf den zweiten aber in seinen Konsequenzen so unabsehbar, dass es an die Tragfähigkeit philosophischer Grundbegriffe rührt. Denn die Möglichkeit, wider besseres Wissen zu handeln, scheint in einem eklatanten Widerspruch zu unserem Selbstverständnis als rational und intentional Handelnde zu stehen.“ (Rebentisch & Setton 2009, 13)

Paradoxien und allgemein Widersprüche haben die Eigenschaft, bewährte Prämissen und Annahmen unseres Selbst- und Weltverständnisses in markanter Weise auf die Probe zu stellen. Das Phänomen der Willensschwäche präsentiert uns mit solch einer Paradoxie in Gestalt einer Antinomie zweier widerstrebender Intuitionen über die Natur unserer praktischen Rationalität, unserer Fähigkeit, unser Handeln durch Gründe leiten und bestimmen zu lassen. Einerseits haben wir die starke Intuition, dass unsere Überlegungen über unsere Gründe, etwas zu tun, uns dazu befähigen, auf Basis dieser Überlegungen unser Handeln zu bestimmen. Wenn ich darüber nachdenke, auf welche Weise ich am besten meinen Urlaub verbringen *könnte*, so hat dieses Nachdenken nur dann für mich Sinn, wenn ich damit die Hoffnung verbinde, auf Basis meiner Überlegungen bessere Entscheidungen darüber zu treffen, wie ich meinen Urlaub verbringen *werde*. Dies impliziert aber, dass ich auch nach Abschluss meiner Überlegungen mein Handeln wirklich an diesen Überlegungen ausrichten kann, d.h. entsprechend meinen Urlaub verbringe. Das Nachdenken über meine Gründe zu handeln wäre nutzlos im Sinne einer Gestaltung meiner Handlungen, wenn es bloß retrospektives Erforschen meiner Gründe involvierte und nicht auch prospektives Erschließen neuer Handlungsoptionen ermöglichte. Letzteres kann es aber nur leisten, wenn zugleich damit die Fähigkeit verbunden wäre, wirklich entsprechend meiner Überlegungen zu handeln. Ansonsten wären solche Überlegungen, die niemals praktisch wirksam wären, in Wahrheit keine praktischen Überlegungen, sondern bloße Theorien der Praxis, d.h. sie wären nur praktisch in ihrem Gegenstand, nicht in ihrem Ergebnis.<sup>1</sup>

Auf der anderen Seite haben wir starke Intuitionen, oft verbunden mit eigenen Erfahrungen, dass unsere Überlegungen über unsere Gründe, etwas zu tun, manchmal ungenügend sind um unsere Handlungen zu bestimmen. Trotz unserer Überlegungen und wider bessere, reflexiv erschlossene Einsicht, handeln wir anders als es unsere Überlegungen empfehlen. Wir tun nicht das, was wir

---

<sup>1</sup> „Reasoning that stops at conditional judgements [...] is practical only in its subject, not in its issue.“ (Davidson 1969a, 39)

reflexiv als das Beste oder Bessere zu tun ansehen. Unser eigenes Handeln erscheint uns in Folge als unverständlich, da wir angesichts unserer Überlegungen, die gegen unsere Handlung sprechen, keinen Grund sehen, weshalb wir so gehandelt haben. Wir verstehen uns selbst nicht, und diese Unverständlichkeit führt eine Fremdheit in uns ein, die wir schwer mit unserem Selbstbild als rational agierende Personen vereinbaren können. Dabei haben wir in diesen Fällen jedoch nicht das Gefühl, bloße Marionetten fremder Mächte zu sein, die durch uns hindurch handeln, sondern selbst als Urheber der Handlungen zu agieren. Wie kann es aber sein, dass wir normalerweise unsere Handlungen an unsere Überlegungen anschließen, dass diese natürlicherweise aus ihnen zu folgen scheinen, und manchmal nicht?

Das Ziel dieser Arbeit ist es, eine Deutung dieses Selbstbildes anzubieten, in der eine solche Integration möglich ist. Die eigene Willensschwäche wird dabei als notwendiges Nebenprodukt unserer Fähigkeit zur Selbstbestimmung verstanden: In dem Maße, wie sich uns durch unsere Überlegungen über unsere Handlungsgründe neue Handlungsoptionen erschließen und mögliche Rechtfertigungen dieser Handlungen anbieten, schaffen sie auch Raum für fehlerhafte Weisen der Selbstbestimmung. Der Ursprung dieser Fehlleistung liegt in der Tatsache, dass neben dem System reflexiver Selbstbestimmung ein System natürlicher Handlungskontrolle und -ausrichtung besteht, das Strebungen für Handlungen umfasst und automatisierte Handlungssequenzen bereitstellt. Neben rationalen, auf Werte bezogenen Wünschen, die auf das aus der Sicht der Person Gute gerichtet sind, existieren präreflexive Formen von Wünschen, auf Emotionen oder Trieben basierende Strebungen, welche nicht wertbezogen sind und daher nicht unbedingt auf das aus der Sicht des Handelnden Gute gerichtet sein müssen. Die Person ist nur ein *Projekt* der idealen Integration solcher gegenläufigen Strebungen und faktisch immer nur partiell ‚integriert‘. Dies legt nahe, dass ohne ein gewisses Maß an Spaltung bzw. Konflikt ein Phänomen wie Willensschwäche nicht verständlich ist.

Auf einer tieferen Ebene konfrontiert uns Willensschwäche auch mit scheinbar widerstreitenden Auffassungen von uns selbst als *frei* handelnde Akteure. Etwas formaler ausgedrückt können wir diesen Widerstreit als Freiheitsantinomie ausdrücken. Wie bei einer Antinomie üblich, erscheinen beide Thesen für sich plausibel zu sein, während sie beide zusammen nicht wahr sein können.

*Thesis:* Frei zu handeln bedeutet für uns, nach Gründen zu handeln, und wir stellen Überlegungen zu dem Zwecke an, die besten Gründe für unser Handeln zu finden. Unsere Handlungen sind demnach dann frei, wenn sie sich nach den besten Gründen richten, die wir in unseren Überlegungen auch als solche auszeichnen können. Es gibt aber keinen Grund, gegen unsere besten Gründe zu handeln. Also handeln wir unfrei, wenn wir gegen unsere besten Gründe

handeln, da wir keinen Grund haben, dies zu tun. Mithin ist das Handeln gegen die von uns als beste erachteten Gründe unfrei.

*Antithesis:* Frei zu handeln bedeutet, nicht gezwungen zu sein, auf eine bestimmte Weise zu handeln. Gründe können unser Handeln nur dann bestimmen, wenn wir uns von ihnen bestimmen lassen. Wenn wir demnach nicht gegen unsere in unseren Überlegungen als beste erachteten Gründe handeln könnten, würden uns unsere besten Gründe zwingen, auf eine bestimmte Weise zu handeln. Gegen unsere besten Gründe handeln zu können heißt also auch, frei handeln zu können.

Unsere Intuitionen über unsere Freiheit führen uns also in widerstreitende Ansichten über die Natur unserer Freiheit: Auf der einen Seite die Ansicht von Freiheit als rationale Selbstbestimmung, auf der anderen Seite die der Freiheit als Ungehindertsein von allen äußeren und inneren Zwängen. Nun liegt der Einwand nahe, die Freiheit, die in der Antithesis ausgedrückt wird, sei eigentlich eine Form von Unfreiheit. Gegen unsere besten Gründe zu handeln sei nicht nur irrational, sondern auch Ausdruck einer Unfreiheit, das eigene Handeln nicht an den besten Gründen auszurichten, die man für sein Handeln anführen kann. Eine solche Freiheit wäre demnach nicht nur eine Scheinfreiheit, sie wäre darüber hinaus auch gar nicht wünschenswert, da sie gerade verunmöglichen würde, was sie beansprucht auszudrücken: Freiheit im Sinne von Ungehindertsein, als Bestimmung des Selbst über sich selbst, oder anders ausgedrückt als Autonomie. Andererseits kann eine Vertreterin der Antithesis geltend machen, dass wohl ein wesentliches Moment von Freiheit die Abwesenheit von Zwang sei. Wenn unser Verhältnis zu unseren Gründen nun ein solches Zwangsverhältnis impliziere, dann sei dies ebenso abzulehnen wie alle anderen Formen von innerem oder äußerem Zwang. Ein Handeln, das sich mechanisch nach den besten Gründen ausrichte, sei kein freies Handeln, da es den handelnden Akteur als bloßes Durchgangsstadium einer Welt von Gründen auszeichne, denen er zu folgen habe. Die Nichtdeterminiertheit des eigenen Handelns setze dagegen voraus, dass nichts anderes als der Akteur bestimme, nach welchen Gründen er sein Handeln ausrichte, welche Gründe damit wirksam sind und welche nicht. Wenn nun Gründe von selbst wirken würden, dann untergrabe dies die Freiheit des Akteurs, sein Handeln selbst zu bestimmen, da nicht er, sondern die Gründe über sein Handeln entschieden. Damit wäre das Handeln wider die besten Gründe Ausdruck unserer Freiheit, und nicht Kennzeichen von Unfreiheit.

Diese Arbeit wird die grundlegende philosophische Antinomie um die menschliche Freiheit nicht zu lösen versuchen, sondern auf einer etwas höheren Ebene der Betrachtung ansetzen. Einander widerstreitende Positionen wie die eben skizzierten haben nämlich die Eigenschaft, fruchtbare Diskussionen über grundlegende Themen auszulösen. Im Falle von Willensschwäche sind dies

Diskussionen über das Verhältnis von Evaluation und Motivation, die Natur intentionalen Handelns, die Zuschreibung von Verantwortung für unsere Handlungen und unser Verständnis von (praktischer) Rationalität und Irrationalität.

Ein zentraler Punkt der philosophischen Beschäftigung mit Willensschwäche bildet das Verhältnis von Evaluation und Motivation, vor allem in der Frage nach dem Zusammenhang von evaluativen Urteilen und Wünschen. Das Urteil der willensschwachen Person über ihre Handlungsgründe wird dabei als evaluatives oder normatives Urteil interpretiert, schließlich ist es eine Stellungnahme des Akteurs zur Güte seiner Gründe zu handeln. Wenn nun evaluative Urteile als in einem strikten Zusammenhang mit den Wünschen des Akteurs stehend gedacht werden, so ergeben sich Schwierigkeiten mit der Erläuterung willensschwacher Handlungen. Obwohl nämlich die willensschwache Person das evaluative Urteil fällt, hat sie nicht den entsprechend starken Wunsch, die in ihren Augen bessere Handlung auch zu vollziehen, da sie schließlich die willensschwache Alternative vollzieht und immer die Handlung vollzogen wird, zu der jemand am meisten motiviert ist. Am meisten motiviert für eine Handlung zu sein bedeutet aber, den stärksten Wunsch bzw. die stärkste Strebung für diese Handlung zu haben. Daher kann das evaluative Urteil über eine Handlung nicht in einem strikten Zusammenhang zum Wunsch, diese Handlung auch auszuführen, stehen, da in Fällen von Willensschwäche die evaluative Bewertung der Handlung und die Motivation für die Handlung auseinanderklaffen. Dabei böte ein solch strikter Zusammenhang zwischen evaluativen Bewertungen von Handlungsoptionen und motivationalen Zuständen des Akteurs eine einfache Erklärung dafür, wie evaluative Urteile praktisch wirksam sein können, d.h. wie Akteure motiviert sein können, normativen Anforderungen nachzukommen. Mit dem Fällen eines evaluativen Urteils, gemäß diesen sogenannten *internalistischen* Theorien, ist die Motivation, d.h. der Wunsch, entsprechend zu handeln, notwendigerweise verbunden. Willensschwäche stellt diesen notwendigen Zusammenhang gerade in Frage und fungiert somit als eine Art Gegenbeispiel für solche Theorien. Die Reaktion auf die Herausforderung durch Willensschwäche fällt verschieden aus. Zumeist wird versucht, Fälle willensschwachen Handelns auf alternative Weise zu beschreiben, die kompatibel ist mit dem angenommenen engen Zusammenhang von Evaluation und Motivation. Bereits Platons Diskussion von Willensschwäche zeigt diese Form des Umgangs. Platon leugnet nicht einfach das Phänomen der Willensschwäche, er leugnet lediglich das Phänomen *in der Beschreibung* als Handlung eines Akteurs gegen sein besseres Wissen über das Gute. Der Willensschwache nach Platon verfügt überhaupt nicht über das relevante Wissen über das Gute und kann daher nicht als jemand beschrieben werden, der gegen ein solches Wissen handelt. Fälle von Willensschwäche sind demnach nicht als Fälle des vorsätzlichen Handelns gegen das bessere Wissen über das Gute zu beschreiben, sondern als Fälle des Nichtwissens bzw.

der Täuschung. Was für die eine Theorie als Leugnung des in Frage stehenden Phänomens gilt, ist für die andere eine Redeskription eines ansonsten unverständlichen Phänomens. Sogenannte *externalistische* Ansätze, welche einen nicht so engen Zusammenhang zwischen Evaluation und Motivation sehen, sind auf den ersten Blick besser geeignet, mit der Herausforderung durch Willensschwäche umzugehen. Wenn Evaluation und Motivation prinzipiell auseinandergehen können, so ist Willensschwäche einfach eines der Fälle, wo genau dieses geschieht: Ein Akteur bewertet eine Handlungsoption in einer Weise die nicht in Deckung gelangt zu seiner Motivation, diese Handlung auch zu vollziehen. Die Schwierigkeit dieser Ansätze besteht nun darin, das Verhältnis von Evaluation und Motivation genauer zu bestimmen und Faktoren auszumachen, welche in Fällen wie Willensschwäche zu einem Auseinanderklaffen der beiden führen. Weshalb ist denn der Akteur nicht motiviert genug, die als besser erachtete Handlung zu vollziehen, sondern eine andere? Basiert sein evaluatives Urteil auf einer Fehleinschätzung seiner tatsächlich vorhandenen Wünsche? Oder verhindern andere Faktoren, dass seine Wünsche mit seinen evaluativen Einschätzungen zur Deckung gelangen? Was hätte der Akteur tun können (oder sollen), um seine Wünsche seinen evaluativen Urteilen anzupassen? Dies alles sind sehr weitreichende und komplexe Fragen, die durch Willensschwäche aufgeworfen werden. Eine mögliche Antwort in der Debatte besteht darin, die Anpassung der Wünsche des Akteurs an seine evaluativen Urteile selbst als ideale Forderung an den Akteur zu interpretieren und sie zugleich damit als Idealform rationalen Handelns zu kennzeichnen. Rationale Akteure sind definiert dadurch das zu tun, was sie als in einer Situation beste Handlungsoption auszeichnen. Der Fehler des Willensschwachen bestehe gerade darin, dieser normativen Forderung nicht nachzukommen, obwohl er prinzipiell die Fähigkeit hätte, dieses zu tun. Ein Problem mit dieser Antwort besteht darin, dass sie lediglich die durch Willensschwäche gestellte Frage reformuliert: Dass der Willensschwache nicht rational gehandelt habe, sei ohnehin klar, und dass er daher seine Wünsche in Einklang mit seinen evaluativen Urteilen bringen sollte, ebenso. Die Frage war doch gerade, weshalb er dies nicht tut, und was er hätte tun können, um rational zu handeln. Ebenso wird gegen diese Art der Erklärung von Willensschwäche der Vorwurf erhoben, eine bloße Scheinerklärung zu liefern: der Willensschwache handelt willensschwach, weil er willensschwach ist. Dies wirft die Frage auf, was denn eine befriedigende Erklärung von Willensschwäche überhaupt sein könnte, die nicht in dieser offensichtlichen Weise zirkulär wäre.

In dieser Untersuchung wird eine Form des deskriptiven Externalismus und normativen Internalismus vertreten: Evaluation und Motivation sind demnach nicht begrifflich notwendig miteinander verbunden, wobei die Motivation nur eine Eigenschaft des evaluativen Urteils wäre, sondern sie sind auf den evaluativen Urteilen externe Quellen der Motivation, Strebungen und Wünsche, angewiesen. Nur ein Teil von Wünschen, wertbezogene und damit rationale Wünsche,

sind mit evaluativen Einstellungen verbunden. Wenn eine Person in diesem Sinne etwas wünscht, dann findet sie das Gewünschte in einer gewissen Hinsicht auch wünschenswert, mithin wertvoll. Neben diesen als rational charakterisierten Wünschen gibt es jedoch eine Vielzahl an nichtrationalen Strebungen und affektiven Zuständen, denen nicht notwendigerweise ein Werturteil entspricht. Rationale Wünsche sind zudem nicht identisch mit reflexiven Werturteilen, weswegen auch Werturteile nicht einfach identisch mit oder bloßer Ausdruck von rationalen Wünschen sind. Es bestehen demnach mehrere Lücken: Einerseits Lücken zwischen unseren reflexiven Werturteilen und unseren rationalen Wünschen, andererseits zwischen unseren unreflektierten affektiven Zuständen und Strebungen und unseren rationalen Wünschen. Dennoch besteht ein normativer interner Zusammenhang zwischen Evaluation und Motivation, eine rationale Anforderung an Akteure, ihre motivationalen Zustände mit ihren evaluativen Überzeugungen in Einklang zu bringen.

Für die Diskussion über die Natur intentionalen Handelns spielt Willensschwäche ebenfalls die Rolle eines oftmals gebrauchten Gegenbeispiels. Für *kausale* Handlungstheorien sind intentionale Handlungen die kausale Folge von Gründen, welche wiederum aus Wunsch-Überzeugungs-Paaren der Akteure bestehen. Eine Handlung ist dann intentional, wenn sie durch ein entsprechendes Wunsch-Überzeugungs-Paar (auf die richtige Weise) verursacht wird. Der Verweis auf Gründe des Akteurs erfüllt dabei die Doppelfunktion einer Rationalisierung und kausalen Erklärung der Handlung. Der Grund rationalisiert die entsprechende Handlung, indem er diese rechtfertigt, und erklärt zugleich das Auftreten der Handlung, indem er als Ursache der Handlung angesehen wird. Dabei stellt Willensschwäche die plausible Parallelisierung von kausaler und normativer Kraft – je besser die Gründe, desto kausal wirksamer – in Frage: Wenn willensschwache Personen gegen ihre besten Gründe handeln, und ihre Handlungen dabei intentional sind, dann ist diese Parallelität durchbrochen. Diese Art von Theorien sind demnach bestrebt zu zeigen, wie sich willensschwache Handlungen in ihr allgemeines Schema zur Erklärung und Explikation intentionalen Handelns einpassen lassen, oder sie streiten die Intentionalität willensschwacher Handlungen gänzlich ab. Für *teleologische* Handlungstheorien scheint dagegen Willensschwäche ein willkommenes Beispiel zu sein, um die Adäquatheit ihrer Auffassungen auch an Fällen intentionalen Handelns zu bewähren, die vom Normalfall abweichen. Handlungen, so eine mögliche Erläuterung, seien eben keine bloß kausalen Folgen von spezifischen Ursachen, und deshalb nur einem eigenen Schema der Erklärung, teleologischen Handlungserklärungen nach Vorbild des praktischen Syllogismus, zugänglich. Dagegen kann schnell eingewendet werden, dass Erklärungen nach dem Muster des praktischen Syllogismus prinzipiell nur geeignet seien, rationale Formen des Handelns zu erklären. Die Rationalitätsannahme sei gewissermaßen in die Struktur der Erklärung eingebaut, weshalb sie

nichts zur Erklärung irrationalen Handelns beitrage. Letzteres sei einer rationalisierenden Handlungserklärung per se nicht zugänglich, außer man wollte behaupten, dass willensschwache Handlungen rational seien. Dies ist zwar intensiv diskutiert worden, jedoch nur mit dem Vorbehalt, dass einige willensschwache Handlungen rational sein könnten, nicht jedoch alle. Beide Arten von Handlungstheorien drohen an der Erklärung von willensschwachem Handeln zu scheitern: Kausale Handlungstheorien an der Frage nach der adäquaten Ursache willensschwachen Handelns, teleologische Handlungstheorien an dessen irrationalem Charakter.

In dieser Untersuchung erläutere ich, in welchem Sinne willensschwache Handlungen als intentional aufzufassen sind. In der Diskussion um kausale vs. teleologische Handlungstheorien wird letztere verteidigt, auch wenn Willensschwäche kein schlagendes Argument gegen kausale Handlungstheorien bietet aufgrund seines generell prekären Status. Doch auch abseits von Willensschwäche bieten Argumente wie abweichende Kausalketten und die Frage nach dem Verhältnis des Akteurs zu seinen Handlungen gute Argumente gegen kausale Handlungstheorien. Abweichende Kausalketten zeigen, dass spezifische Ursachen alleine noch nicht die Intentionalität von Handlungen sicherstellen. Die Rolle von Akteuren an ihren Handlungen wird in kausalen Handlungstheorien generell unklar konzeptualisiert: Es ist nicht klar, was ein Akteur eigentlich tut, wenn er handelt. Dieses Moment des Tuns oder der Aktivität entzieht sich, so die hier vertretene Auffassung, einer kausalen Binnenanalyse. Hier stellt sich auch die Frage, was eine philosophische Handlungstheorie, die eine bestimmte Auffassung von Freiheit vertritt, überhaupt an Erklärungskraft zugemutet werden sollte. Wenn Gründe nicht von sich aus kausal hinreichend sind, sondern immer eines Moments der Anerkennung durch den Akteur bedürfen, um handlungswirksam zu sein, dieses Moment wiederum nicht von anderen kausalen Momenten abhängt, dann wäre es unsinnig, gegen eine solches Konzept den Vorwurf der Nichterklärbarkeit menschlichen Tuns zu erheben. Eine starke Freiheitsauffassung sollte hier gelassen zugeben, dass Freiheit letztlich nicht erklärbar ist, wenn mit einer Erklärung die Rückführung auf kausal hinreichende Gründe verstanden wird. So wird sich auch Willensschwäche letztlich einer gänzlich kausalen Erklärung entziehen, wenn in ihr dieses Freiheitsmoment enthalten ist. Letzteres spricht jedoch nicht gegen die Verstehbarkeit von Willensschwäche als spezifische Weise, die eigene Freiheit in einer Weise auszuüben, welche die genannten strukturellen Momente aufweist.

Ein weiteres wesentliches Moment stellt die Frage nach der Zuschreibung von Verantwortung für Willensschwäche und willensschwache Handlungen dar. Eine oftmals vorgebrachte Argumentationsfigur in diesem Zusammenhang verläuft dabei folgendermaßen: Personen seien verantwortlich für ihre Willensschwäche und ihre willensschwachen Handlungen. Da Verantwortung prinzipiell Freiheit involviere, ausgedrückt durch das sogenannte Prinzip

alternativer Möglichkeiten,<sup>2</sup> schliesse dies aus, willensschwache Handlungen und Willensschwäche als Fälle physischen oder psychischen Zwangs aufzufassen. Die Möglichkeit und die Art und Weise der Unterscheidung willensschwacher Handlungen von erzwungenen Handlungen bildet daher eines der wesentlichen Punkte der Diskussion. Hierzu werde ich auf den Ansatz von Gary Watson zurückgreifen und ihn ausbauen. Dieser hat die übliche Weise, willensschwache Handlungen von Zwangshandlungen zu unterscheiden, in Zweifel gezogen.<sup>3</sup> Nach Watson ist der Willensschwache nicht frei, eine andere Handlung als die willensschwache zu vollziehen, aber er ist dennoch für seine Tat verantwortlich, da er es verabsäumt hat, gewisse Fähigkeiten und Dispositionen in genügendem Maße auszubilden, die es ihm ermöglicht hätten, nicht willensschwach zu handeln. Die Zuschreibung von Verantwortung ist in diesen Fällen unabhängig davon, ob in der jeweiligen konkreten Situation die Person anders hätte handeln können, als sie es getan hat. Das Modell der Zuschreibung von Verantwortung für willensschwache Handlungen ist also analog zur Zuschreibung von Verantwortung für fahrlässige Handlungen zu denken.

Die Frage danach, ob und weshalb wir Willensschwäche vermeiden sollen, führt auf die evaluative und normative Dimension von Willensschwäche. Normativ gesehen verstoßen willensschwach Handelnde gegen Anforderungen praktischer Rationalität. Dabei hängt es von unserem Verständnis der Reichweite dieser Anforderungen ab, wie wir die Irrationalität von Willensschwäche verstehen. In einem engen, internen Verständnis begeht die willensschwache Person einen Fehler im Sinne der Inkohärenz mentaler Zustände. In einem externen und weiten Verständnis kann es jedoch sein, dass die willensschwache Handlung zumindest als rationaler erscheint als ihre willensstarke Alternative, wenn auch nicht als völlig vernünftig. Evaluativ betrachtet werde ich die These vertreten, dass Willensschwäche deshalb vermieden werden sollte, da sie tendenziell die Selbstachtung von Personen untergräbt, welche wiederum Bedingung der Möglichkeit für ein gelungenes Leben einer jeden Person ist.

Ich beginne die Untersuchung mit einer Bestimmung des in Frage stehenden Phänomens (*Kapitel I*). Gegenstand der Untersuchung ist hauptsächlich Willensschwäche im Sinne des intentionalen Handelns wider das eigene, bessere Urteil. Da jede Abgrenzung notwendig einen Blick

---

<sup>2</sup> „A dominant role in nearly all recent inquiries into the free-will problem has been played by a principle which I shall call “the principle of alternate possibilities.” This principle states that a person is morally responsible for what he has done only if he could have done otherwise.” (Frankfurt 1969, 1)

<sup>3</sup> „To take seriously the possibility of acting contrary to one's better judgment is at the same time to raise problems about the distinction between weakness and compulsion. I have argued that the common view, according to which the differentiating feature is that the weak are able to conform their behavior to their practical judgments, is unjustified. Instead, I have proposed that weakness of will involves the failure to develop certain normal capacities of self-control, whereas compulsion involves desires which even the possession of such capacities would not enable one to resist.” (Watson 1977, 339)

auf das dadurch Abgegrenzte impliziert, sollen hier alternative Bestimmungen des Phänomenbestands der Willensschwäche diskutiert werden. In *Kapitel II* stelle ich zwei zentrale Konzeptionen von Willensschwäche, diejenige von Aristoteles und Donald Davidson, dar um in die spezifisch philosophischen Probleme im Zusammenhang mit Willensschwäche einführen. Nicht nur hatten beide Konzeptionen wesentlichen Einfluss auf die philosophische Diskussion, sondern stehen auch paradigmatisch für die moralphilosophische und handlungstheoretischer Perspektive auf Willensschwäche. Die aristotelische Sicht spielt neben dieser problemexpositorischen Funktion außerdem auch eine inhaltliche Rolle für das Verständnis des Verhältnisses von Evaluation und Motivation, während Davidson wiederum immer wieder als Bezugspunkt der Diskussion dienen wird. *Kapitel III* widmet sich dem Problemfeld des Verhältnisses von evaluativen Urteilen und der Motivation, entsprechend der Urteile zu handeln. Hier diskutiere ich verschiedene Theorien, welche dieses Verhältnis näher bestimmen, und deren relative Bedeutung für Willensschwäche. In *Kapitel IV* verfolge ich die Frage nach der Intentionalität willensschwacher Handlungen. Dabei werde ich verschiedene Bedeutungen von Intentionalität voneinander unterscheiden und hinsichtlich ihrer Kompatibilität zu Willensschwäche untersuchen. Hier werde ich auch näher auf exekutive Willensschwäche eingehen. In *Kapitel V* stelle ich die Frage nach der Verantwortung von Personen für ihre Willensschwäche und willensschwachen Handlungen. Dabei wird die These von Gary Watson in ihren Grundsätzen verteidigt. In *Kapitel VI* stelle ich die Frage nach der evaluativen und normativen Bewertung von Willensschwäche. Dabei diskutiere ich verschiedene Konzeptionen praktischer Rationalität und praktischer Überlegung, sowie die Rolle von Selbstrespekt für die evaluative Bewertung von Willensschwäche. In *Kapitel VII* trage ich nochmals wesentliche Punkte der Untersuchung kompakt zusammen und biete eine Deutung von Willensschwäche als Kehrseite unserer Fähigkeit zur reflexiven Selbstbestimmung an.

# I Phänomen und Begriff der Willensschwäche

„It is useful to distinguish between (i) giving an account of mind and value that *leaves room* for weakness of will, (ii) *describing* or *anatomizing* weakness of will in convincing detail, and (iii) *explaining* weakness of will.“  
(Wiggins 1978, 275)

Dieses Kapitel dient dazu, mit dem Phänomen der Willensschwäche vertraut zu machen und es näher zu bestimmen. Das Zitat von David Wiggins macht deutlich, dass das Ziel der Beschäftigung mit Willensschwäche je nach Zugriff auf das Phänomen verschieden ist. Während die eigene Willensschwäche z.B. als praktisches Problem der eigenen Lebensführung angesehen wird, das man irgendwie zu lösen versucht, interessiert philosophische Ansätze vor allem die begriffliche Ordnung, in die sich das Phänomen möglichst konsistent einfügen lässt. Ich beginne mit einem kurzen Überblick über mögliche Zugangsweisen zum Phänomen der Willensschwäche (1) und stelle in Folge zwei Typologien von Willensschwäche vor (2). Nach einer Diskussion des Verhältnisses der Begriffe ‚Akrasia‘ und ‚Willensschwäche‘ (3) sowie der Beschäftigung mit der Frage, ob Willensschwäche auch im Bereich theoretischer Rationalität möglich ist (4), unterscheide ich zwei verschiedene Arten von Willensschwäche, *motivationale* und *exekutive* Willensschwäche (5). Diese werde ich auch gegenüber andere, benachbarte Phänomene abgrenzen.

## I.1 Willensschwäche im Alltag, in der Kunst und als Gegenstand der Forschung

Wie und wo begegnen wir Willensschwäche? Wie ist das Phänomen in unserem Alltag verankert, wie erfahren wir Willensschwäche, oder phänomenologisch gesprochen: wie ist Willensschwäche in der Erfahrung gegeben? Die Antwort auf diese Frage hängt davon ab, welche Erfahrung im Blick ist, da die Weisen der Erfahrung eines Phänomens untrennbar mit der Eigenart des gegebenen Phänomens verknüpft sind. Im Alltag erfahren wir Willensschwäche auf zwei Weisen: als unsere eigene Willensschwäche und als diejenige der anderen. Das Problem mit diesen Erfahrungen ist nicht, dass sie nicht vorhanden wären, sondern dass sie mit vielerlei anderen Phänomenen in enger Verbindung stehen und nicht klar ist, wie und wo die Grenzen zwischen diesen Phänomenen verlaufen. Ist jemand willensschwach, mutlos oder vielleicht beides? Bin ich selbst willensschwach, wenn ich mir wieder einmal vornehme, mehr Sport zu treiben und es nicht tue, oder täusche ich mich selbst, wenn ich mir dies immer wieder vornehme und glaube, es wirklich ernst damit zu meinen? Für die philosophische Argumentation ist der Verweis auf die eigene Erfahrung der eigenen oder fremden Willensschwäche daher immer *problematisch*: der Bericht einer unreflektierten Erfahrung kann der einen Person eingängig, der anderen unplausibel oder fremd vorkommen. In vielen philosophischen Diskussionen wird auf solche Berichte

fingierter alltäglicher Erfahrungen verwiesen um für oder gegen die Plausibilität philosophischer Positionen zu argumentieren. Leider ist der Wert dieser Berichte nur begrenzt. Auch wenn die alltägliche Erfahrung letztlich das Fundament und der Horizont aller anderen Erfahrungen und deren Geltungsansprüche abgibt, kann nicht so einfach *in* die Erfahrung hineingeblickt werden, um bestimmte Phänomene besser zu verstehen bzw. begrifflich zu fassen.

Der vielfach bemerkte prekäre Charakter des Phänomens der Willensschwäche bedingt eine Gegenüberstellung zweier symmetrischer Argumentationsmuster, die beide mit gewissem Recht vertreten werden können und sich dennoch in ihren Konklusionen einander widersprechen. Willensschwäche Skepsis schließt auf Basis von handlungstheoretischen und metaethische Überlegungen darauf, dass es Willensschwäche, beschrieben in einer bestimmten Weise, nicht geben kann. Den umgekehrten Weg beschreiten Dogmatiker, welche auf Basis von bestimmten Erfahrungen von Willensschwäche, die sie in einer bestimmten Weise beschreiben, gegen oder für bestimmte Handlungstheorien und metaethische Überzeugungen argumentieren.

**Willensschwäche-Skepsis (top-down-approach)**

Prämissen: Handlungstheoretische und/oder moralphilosophische Theorien

Konklusion: Willensschwäche beschrieben als X ist unmöglich

**Willensschwäche-Dogmatik (bottom-up-approach)**

Prämissen: Erfahrungen von Willensschwäche beschrieben als X

Konklusion: Handlungstheorien und/oder moralphilosophische Theorien müssen so beschaffen sein, dass sie X als möglich ausweisen

In einer solchen Situation ist es unwahrscheinlich, dass eine Seite ein Argument vorbringen kann, das die andere Seite dazu bringt, von ihrer Position abzuweichen. Dementsprechend unternehme ich nicht einen solchen Versuch, sondern zeige a) welche argumentativen Verschränkungen zwischen verschiedenen Beschreibungen von Willensschwäche und handlungstheoretischen wie moralphilosophischen Annahmen bestehen und b) wie eine besonders anspruchsvolle Form der Beschreibung von Willensschwäche kohärent mit plausiblen handlungstheoretischen und metaethischen Grundannahmen gedacht werden kann.

Die Kunst ist in vielerlei Hinsicht besser geeignet als unreflektierte alltägliche Erfahrung, um mit dem Phänomen der Willensschwäche bekannt zu machen, da sie (oft) auf Allgemeines in der Einzelheit abzielt, auf die Sichtbarmachung allgemeiner Typen und Strukturen in konkret erlebbaren Erfahrungen. Als Beispiele für Willensschwäche wird für die Literatur immer wieder

gerne auf die Figur der *Medea* in der Tragödie von Euripides verwiesen.<sup>4</sup> Medea sinnt auf Rache und plant neben ihren Ehemann Jason auch ihre eigenen Kinder zu ermorden. Der daraus resultierende innere Widerstreit wird vielfach als Konflikt zwischen Vernunft und irrationalem Affekt interpretiert: Medea tötet ihre Kinder wider bessere Einsicht aus irrationalem, übermäßigen Zorn – ein Fall von Willensschwäche. Der wissenschaftliche Zugriff auf das Phänomen der Willensschwäche wiederum ist vielfältig entsprechend der Vielfalt an wissenschaftlichen Gebieten, Theorien, Methoden und Erfahrungsweisen. Eine besondere Rolle spielt Willensschwäche sicherlich für Wissenschaften, in denen der Begriff des rationalen Handelns wichtig ist. Innerhalb der Psychologie betrifft Willensschwäche vorwiegend motivationspsychologische Fragestellungen. Dies betreffen Phänomene wie Selbstkontrolle, Selbstregulation und Affektkontrolle. Natürlich spielt Willensschwäche auch eine Rolle in Bereichen der Suchtforschung. Innerhalb der Ökonomie wird Willensschwäche vor allem als dynamische Inkonsistenz behandelt, als zeitlich einander widersprechende Präferenzstrukturen, ausgelöst durch irrationale Weisen der Diskontierung zukünftigen Nutzens.

## **I.2 Formen und Typologien von Willensschwäche**

In der philosophischen Diskussion taucht der Begriff der Willensschwäche in der Antike als ‚Akrasia‘ erstmals auf. Der griechische Begriff ‚Akrasia‘ ist gebildet aus dem Alpha privativum (als Ausdruck der Negation) und ‚kratos‘ (Macht, Kraft) und bedeutet demnach angewendet auf Akteure ursprünglich so viel wie ‚fehlende Selbstkontrolle‘ oder ‚Unbeherrschtheit‘. Für die philosophische Debatte wurde der Begriff von Platon und Aristoteles zur Beschreibung von Personen eingeführt, die gegen ihr besseres Wissen, was gut sei, handeln. Die traditionelle Debatte drehte sich vorwiegend um die Fragen, worin die Unbeherrschtheit genau bestehe, auf was sie bezogen und ob und wie sie überhaupt möglich sei. Die Untersuchungen zu Akrasia waren in allgemeine moralphilosophische Theorien eingebettet, deren Hauptziel die Frage nach dem guten Leben bildete. In modernen philosophischen Diskussionen wurde das Phänomen der Akrasia aus diesem moralphilosophischen Kontext gelöst und als allgemeines Problem der Handlungstheorie betrachtet: Wenn zu handeln heißt, etwas aus einem bestimmten Grund zu tun, dann stellt sich die Frage, wie jemand ungezwungen gegen seine Überlegungen, was am besten zu tun sei, handeln kann. Die moderne Debatte hat jedoch auch die Diskussion um Akrasia mit Untersuchungen zu Willensschwäche in Verbindung gebracht und zumeist diese miteinander identifiziert. Das Verhältnis von Akrasia und Willensschwäche ist umstritten, auch was die Interpretation der alltäglichen Verwendung des Begriffs ‚Willensschwäche‘ betrifft.<sup>5</sup>

---

<sup>4</sup> Vgl. Müller 2009, Kap. II.1.2.

<sup>5</sup> Vgl. Mele 2010a; Holton 1999; Dodd 2009.

Einige Autoren vertreten die Ansicht, dass Willensschwäche streng von Akrasia zu unterscheiden sei: Während Akrasia im Handeln wider das bessere Urteil oder Wissen des Akteurs bestehe, sei Willensschwäche die irrationale Änderung oder Revision von Entscheidungen oder Absichten. Während der Akratiker sich angesichts seines besseren Wissens für die schlechtere Handlung entscheide, scheitere der Willensschwache an der adäquaten Umsetzung einmal getroffener Entscheidungen. Ich unterscheide diesen beiden Formen auch, doch nicht über die aus meiner Sicht etwas willkürliche Terminologie von ‚Akrasia‘ und ‚Willensschwäche‘, sondern unter den Bezeichnungen *motivationale* und *exekutive* Willensschwäche.

Das Phänomen der Willensschwäche wurde auch im Kontext semantischer Analysen normativer Begriffe diskutiert, v.a. im Zusammenhang mit Richard Hares Präskriptivismus.<sup>6</sup> Ein Akteur kann demnach bestimmte normative Ausdrücke wie ‚sollen‘ oder ‚müssen‘ nur dann aufrichtig verwenden, wenn sein Handeln auch den durch sie ausgedrückten Vorschriften folgt. Der Fall, dass jemand aufrichtig meine, er solle etwas tun, und es dennoch nicht tue, obwohl er es tun könne, sei daher gar nicht möglich. Der Willensschwache sei psychisch unfähig, seinen eigenen Vorstellungen darüber, was er tun sollte, nachzukommen, oder er verwende dabei die zentralen normativen Begriffe nicht in ihrem ursprünglich präskriptiven Sinn, sondern auf abgeschwächte Weise. Die Diskussion um Willensschwäche wurde auch oft in Bezug gebracht zur Diskussion um das Phänomen der Selbsttäuschung. Einige Theorien gehen davon aus, dass Willensschwäche im Grunde eine Form der Selbsttäuschung des Akteurs über seine wahren Motive darstelle.<sup>7</sup> Im Folgenden gilt es zu klären, was sinnvollerweise als Willensschwäche bezeichnet werden sollte bzw. welche Phänomene unter diesen Begriff fallen sollen. Die philosophische Debatte über die Frage, was alles unter ‚Willensschwäche‘ fällt, ist, wie die kurzen Andeutungen oben bereits zeigen sollten, äußerst komplex und schwierig nachzuzeichnen. Generell lassen sich zwei Richtungen voneinander unterscheiden:

1) Positionen, welche auf einen sehr begrenzten Begriff von Willensschwäche als ‚Handeln wider das bessere Urteil oder Wissen‘ abzielen. Handeln wider das bessere Wissen ist dabei enger als Handeln wider das bessere Urteil, da Wissen notwendigerweise die Wahrheit des Gewussten impliziert, während ein bloßes Urteil auch falsch sein könnte. Zu dieser Gruppe gehören die traditionellen Beiträge von Platon und Aristoteles, aber auch die von Donald Davidson. Dazu gehören auch Positionen, welche Fälle von Akrasia von Fällen von Willensschwäche und anderen Formen praktischer und theoretischer Irrationalität abgrenzen wollen.

---

<sup>6</sup> Vgl. Hare 1952; Kap. III.3.1.

<sup>7</sup> Vgl. Wolf 1985; Schälike 2004; Kap. III.1

2) Positionen, welche einen weiten Begriff von Willensschwäche vertreten, also dessen Anwendung nicht auf Fälle des Handelns wider das bessere Urteil oder Wissen begrenzen, sondern Willensschwäche als Überbegriff von verschiedenen Arten praktischer, manchmal sogar theoretischer Irrationalität, verwenden.

Ich denke, dass die einheitliche Bezeichnung von diversen Phänomenen als ‚Willensschwäche‘ nur dann Sinn macht, wenn klar ist, dass sich diese diversen Phänomene in einer Hinsicht gleichen. Willensschwäche wäre dann ein Überbegriff für eine Familie von Phänomenen, die alle dieselbe Struktur teilen. In der philosophischen Literatur zu Akrasia und Willensschwäche werden jedoch viele verschiedene Phänomene beschrieben, die auf den ersten Blick keine einheitliche Struktur aufweisen, und deshalb auch in der Analyse voneinander getrennt werden sollten. Genannt werden unter anderem: Handeln wider das bessere Urteil/Wissen; vorschnelle oder irrationale Revision von Handlungsvorsätzen; das halbherzige Ausführen von Handlungen; das Fassen schwacher Vorsätze; der Verzicht auf das Fassen von Vorsätzen; die Änderung von Urteilen darüber, was am besten zu tun sei; apathisches und depressives Verhalten; vorschnelles Handeln ohne zu überlegen, wie z.B. Handlungen aus Zorn oder Eifersucht; das Ziehen falscher praktischer Schlüsse, was am besten zu tun sei; die fehlerhafte Interpretation von Handlungssituationen; das Haben unangepasster Emotionen usw. Bereits diese unvollständige Liste zeigt, dass offensichtlich sehr disparate Phänomene untersucht werden, und es scheint unklar, auf welche Weise alle diese Phänomene eine gemeinsame Struktur zeigen, die eine einheitliche Bezeichnung der Phänomene rechtfertigen würde. Ich denke, dass es daher besser ist, an den konkreten Phänomenen zu bleiben und diese bestmöglich zu analysieren. Gleichzeitig kann man diese diversen Phänomene versuchen in einen gemeinsamen Rahmen zu stellen, der eine gewisse Ordnung der Phänomene erlaubt, ohne dass diese Ordnung suggeriert, dass die Phänomene irgendwie gleich strukturiert wären oder dass die Ursachen der Phänomene gleich wären. Diese Ordnung würde davon ausgehen, dass die jeweiligen Phänomene unterschiedliche Fähigkeiten und Eigenschaften von Personen betreffen und an verschiedenen Stellen in einem idealisierten Prozess von der Interpretation der Handlungssituation über Überlegungen über mögliche Handlungsoptionen zur Handlungsausführung auftreten können.

### **1.2.1 Klassifikation von Willensschwäche nach Amélie Rorty**

Amélie Rorty hat in ihrem Essay *Where does the akratic break take place?* (1980a) eine Klassifikation verschiedener Arten von Willensschwäche vorgeschlagen, deren Ordnungsprinzip orientiert ist am sogenannten praktischen Syllogismus. Die verschiedenen Arten von Willensschwäche bzw. Akrasia unterscheiden sich voneinander anhand der Stelle im Prozess vom Überlegen zum

Handeln, an der sie auftreten können. Diesen Prozess unterteilt Rorty in fünf verschiedene Phasen:

- „1. A person's general beliefs about appropriate human aims, what is commanded by God, or required by morality, or conduces to human well-being.
2. A person's commitment to actualize or realize those aims, to instantiate them in his actions, to attempt to be guided by them.
3. A person's interpretation of a particular situation:
  - a. His perceptual categorizations: the structuring of what is salient in a situation, at the center of his focused attention
  - b. His verbal descriptions and characterizations of his situation
  - c. His beliefs about that situation
  - d. His emotional responses, as they affect and are affected by, his perceptions
4. A person's forming an intention to act:
  - a. A person's forming an *intention* to act, where the intention is conditional on the satisfaction of some antecedents in one isolated piece of practical reasoning directed by one end. [...]
  - b. A person's forming a *resolution* to act, where the resolution is not detached from the comparative practical reasoning that has formed priorities. [...]
  - c. A person's *deciding* to act according to his resolution, where the *decision* is detached from comparative practical reasoning in which it was imbedded and from which it was derived. The decision is no longer treated as a conditional.
5. A person's acting according to his decision.” (Rorty 1980a, 230)

Rorty versteht dieses Schema als Ergänzung zum praktischen Syllogismus zurückgehend auf Aristoteles. Stufe 1 umfasst allgemeine evaluative Überzeugungen der Person, zu denen sie sich auf Stufe 2 verpflichten kann, sie auch tatsächlich in Handlungen umzusetzen oder sich von ihnen in ihren Handlungen leiten zu lassen. Eine Person, die es nicht schafft, ihre allgemeinen Überzeugungen über das in einem weiten Sinne Gute als handlungsleitend anzusehen, zeigt *akrasia of direction or aim*, also Willensschwäche der Richtung oder des Zieles. Eine Person, die einzelne Situationen nicht in Übereinstimmung mit allgemeinen Vorsätzen, repräsentiert in 2, interpretiert, zeigt *interpretative akrasia* oder interpretative Willensschwäche. Eine Person, die in ihren konkreten praktischen Überlegungen, was in einer gegebenen Situation zu tun sei, falsche Schlüsse zieht – also 2 und 3 falsch oder nicht adäquat verknüpft – zeigt *akrasia of irrationality* oder irrationale Willensschwäche. Schließlich zeigt eine Person, die gegen ihre konkreten praktischen Überlegungen handelt, *akrasia of character* oder Willensschwäche des Charakters. Auf jeder dieser Stufen können noch feinere und spezifischere Arten von Willensschwäche unterschieden werden, wiederum abhängig von der spezifischen mentalen Leistung, die involviert ist. So könnte man noch akratische Emotionen, Wahrnehmungen, Überzeugungen usw. voneinander unterscheiden. Rorty will mit diesem Schema offensichtlich ein möglichst umfassendes System der Klassifikation verschiedenster Arten praktischer Irrationalität bereitstellen.

Rorty nennt zudem noch zwei Bedingungen, welche gegeben sein müssen, damit wirklich ein Fall von Willensschwäche vorliegt: 1) Die Person muss fähig sein, ihre eigene Willensschwäche zu bemerken.<sup>8</sup> 2) Die Person muss in dem Bereich, in dem sie Willensschwäche zeigt, fähig sein, absichtlich anders zu handeln.<sup>9</sup> Dabei besteht für Rorty ein Zusammenhang zwischen den beiden Bedingungen insofern, als eine Person nur dann fähig ist, absichtlich in dem Bereich, in dem Willensschwäche auftritt, anders zu handeln, wenn sie ihre Willensschwäche auch bemerken kann. Die Person, die willensschwach ist, muss fähig sein, den Verstoß gegen die eigenen Präferenzen *als* einen solchen Verstoß zu bemerken und ihn normalerweise vermeiden zu können. Mit dieser Rückbindung der Willensschwäche an das Vorliegen eines reflexiven Zugriffs seitens des Akratikers taucht als Kern der Willensschwäche wiederum das Handeln wider das bessere Urteil auf.<sup>10</sup> Der Unterschied liegt nur darin, dass hier das Urteil der Handlung nicht notwendigerweise vorhergehen muss, und dass der Bereich, auf den sich Willensschwäche bezieht, nicht nur Handlungen sind, sondern auch psychische Zustände wie Überzeugungen, Emotionen, Wahrnehmungen etc. insofern sie beeinflussbar sind durch absichtliche Handlungen. So könnte jemand willensschwach sein, der aus seiner Sicht für eine Situation unangepasste Emotionen hat, wenn sie direkte Folge von absichtlichen Handlungen oder dispositionale als Reihe von Handlungen und Reaktionen zu analysieren sind.<sup>11</sup>

Es gibt nun mehrere Bedenken gegen solch eine extrem weite Bestimmung von Willensschwäche. *Erstens* ist Willensschwäche in dieser Form nicht unterscheidbar von praktischer Irrationalität überhaupt, solange sie durch den Akteur erkennbar und im Einflussbereich des absichtlichen Handelns seitens des Akteurs liegt. Willensschwäche ist in dieser Form aber auch nicht unterscheidbar von Formen theoretischer Irrationalität, solange diese Bedingungen erfüllt sind. Ein vermeidbarer Rechenfehler aufgrund von mangelnder Konzentration z.B. wäre auch willensschwach, wenn die Person das präferierte Urteil hat, dass sie konzentrierter sein sollte. Hier stellt sich die Frage, ob der praktische Syllogismus als Ausgangspunkt einer solchen Klassifikation überhaupt noch passend ist. *Zweitens* bleibt unklar, wie der Status des präferierten Urteils genauer zu fassen ist. Dies sind anscheinend nicht nur auf Handlungen bezogene Urteile, sondern auch solche, wie zuvor im Rechenbeispiel, über Standards der Wissensgenerierung, der

---

<sup>8</sup> „One of the conditions for akrasia is that the agent be capable of recognizing that he has violated his preferences.” (Rorty 1980a, 244)

<sup>9</sup> „Someone capable of akrasia of a particular kind is capable of voluntary action in that domain: this means that he is in principle the sort of person who could have done otherwise.” (Rorty 1980a, 245)

<sup>10</sup> „If he were incapable of recognizing that his action violates his preferred judgment, then he would not be the sort of person for whom that action is properly voluntary. This condition is a reflexive one. Not only does the akratic agent have preferences, but he must also be capable of recognizing them as *his* preferences. If the akratic action is voluntary, the agent must be capable of avoiding it, not just accidentally avoiding it, but avoiding it because he takes it to be a violation of his preference.” (Rorty 1980a, 245)

<sup>11</sup> Vgl. Rorty 1980a, 237f.

Angemessenheit von Emotionen, der Adäquatheit sprachlicher Beschreibungen von Situationen usw. Im Grunde fallen alle möglichen Arten von Normen darunter, solange sie in Urteilen der Person abgebildet werden können. Es erscheint aber unangemessen zu sein, jede Art des Verhaltens gegen diese Normen als willensschwach zu bezeichnen. Auch bleibt unklar, wann und zu welchem Zeitpunkt der Willensschwache die fraglichen Urteile fällt und seine Willensschwäche bemerkt. *Drittens* haben wir gesehen, dass der Kern der Willensschwäche, das Handeln wider das bessere Urteil, auch in dieser Analyse wieder auftaucht. Eine Analyse dieses Kernes wäre demnach auch für diese weite Bestimmung von Willensschwäche notwendig.

### 1.2.2 Ein interdisziplinäres Rahmenmodell für Willensschwäche

Einen interdisziplinären Zugang zur Bestimmung von Willensschwäche entwerfen Kalis, Mojzisch, Schweizer und Kaiser in ihrem *Weakness of will, akrasia, and the neuropsychiatry of decision making: An interdisciplinary perspective* (2008). Die Bemühung geht dahin, die philosophische, psychologische und psychopathologische Perspektive auf verschiedene Formen praktischer Irrationalität zusammenzubringen. Dabei entwickeln sie ein generelles Schema, das es ermöglichen soll, verschiedene Arten von praktischer Irrationalität – pathologische wie nicht-pathologische Formen – zu unterscheiden und so einer angemesseneren Analyse zuzuführen. Das Schema basiert auf einer Dreiteilung des Entscheidungsprozesses von Akteuren für Handlungen in eine Phase der *Generierung von Handlungsoptionen*, eine der *Abwägung der generierten Optionen* mit abschließender Auswahl einer der Optionen und schließlich eine der *Initiierung dieser Handlungsoption*:

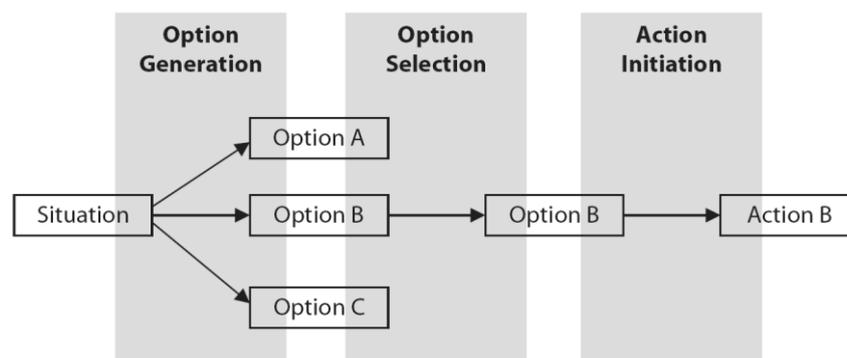
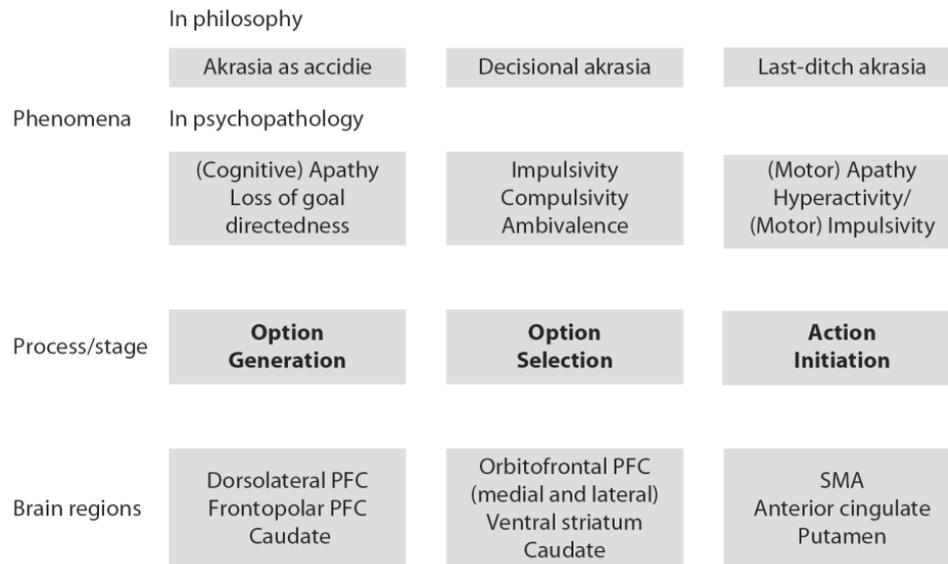


Figure 1. Stages of decision making in our model.

(Kalis & al. 2008, 403)

Diese Phasen werden nun mit philosophischen, psychopathologischen und neurobiologischen Erklärungsmustern und Phänomenbeschreibungen von Abweichungen zum normalen Entscheidungsverlauf korreliert:



**Figure 2. Three levels of explanation are distinguished in our model: (1) phenomena, in philosophical or psychopathological terms; (2) processes or stages of decision making; and (3) associated brain regions.**

(Kalis & al. 2008, 404)

In der *ersten Phase* findet die Generierung von Handlungsoptionen eines Akteurs in einer gegebenen Situation statt. Eine Handlungsoption besteht für die Autoren aus zwei Komponenten:

„First, a cognitive–motor program contains the action, as well as its conditions, on a cognitive–motor level. Second, options also need to have an affective value, in order to motivate the agent to select and initiate an action.” (Kalis & al. 2008, 405)

Handlungsoptionen haben also eine kognitive Komponente, bestehend aus einer Repräsentation des Handlungstypus, und einer affektiven oder motivationalen Komponente, bestehend aus einem Wunsch, bezogen auf einen bestimmten Wert, auf den der Wunsch gerichtet ist und der mit dem Handlungsergebnis verbunden ist. Dysfunktionale Formen der Generierung von Handlungsoptionen betreffen entweder den *Inhalt* der generierten Optionen oder deren *Anzahl*.<sup>12</sup> Hinsichtlich der vorigen Klassifikation von Rorty könnten wir ihre akrasia of direction or aim und interpretative akrasia dieser Phase zuordnen, wobei erstere die Anzahl, letztere den Inhalt der generierten Handlungsoptionen betreffen. Die Anzahl an generierten Handlungsoptionen kann dabei entweder zu klein (Hypogeneration) oder zu groß (Hypergeneration) sein. *Hypogeneration* von Handlungsoptionen wird in Literatur und Philosophie oft als Fälle der generellen Indifferenz, Faulheit, Trägheit oder Apathie beschrieben.<sup>13</sup> Auch eine *Hypergeneration*, eine Überanzahl an Handlungsoptionen, kann einen lähmenden Effekt auf das Handeln einer Person

<sup>12</sup> „Option generation could be dysfunctional because of the form of the process (the number of options generated) or the content of the process (the nature of the generated options).” (Kalis & al. 2008, 407)

<sup>13</sup> „The idea is that sometimes people just do not care about translating their general values into concrete options for action.” (Kalis & al. 2008, 407)

haben, indem sie gar nicht in die Phase gelangt, Handlungsoptionen gegeneinander abzuwägen und so vom Handeln überhaupt absieht. Die Autoren gehen nicht näher auf Fälle ein, in denen dysfunktionale Formen des Inhalts von Handlungsoptionen vorliegen. Ich denke, dass hier Rortys Beschreibungen von interpretative akrasia passen, also Fälle, in denen Personen aufgrund fehlerhafter Wahrnehmungen und Interpretationen ihrer jeweiligen Situation inhaltlich unangepasste Handlungsoptionen generieren.

Die *zweite Phase* umfasst die Abwägung der verschiedenen generierten Handlungsoptionen sowie die Auswahl einer konkreten Option in Gestalt einer Entscheidung. Dies umfasst Rortys Fälle von irrational akrasia und auch ein Großteil der Phänomene, mit der sich die traditionellen und modernen Debatte um Willensschwäche beschäftigen, können dieser Phase zugeordnet werden. Als mögliche dysfunktionale Formen der Auswahl von Optionen nennen die Autoren *Impulsivität*, *Zwanghaftigkeit* und *Ambivalenz*. Bereits Aristoteles hat in seiner Analyse von Akrasia zwischen zwei Formen unterschieden, den Unbeherrschten, die eine Überlegung anstellen und dann entgegen dieser Überlegung handeln, und solchen, die keine Überlegung anstellen und voreilig handeln.<sup>14</sup> In diesem Sinne könnte Voreiligkeit der Impulsivität zugeordnet werden, während die erste Art der Akrasia, die Unbeherrschtheit als Schwäche, keine so direkte Entsprechung findet. Dies ist deswegen so, da sie den Inhalt der Überlegungen betrifft, nicht nur deren Form. Obwohl jemand in seiner Überlegung eine Handlungsoption präferiert, entscheidet er sich für eine andere Handlung. *Impulsivität* zeigt sich z.B. in Verhaltensweisen wie der fehlerhaften Diskontierung zeitlich naher und ferner Gewinne.<sup>15</sup> *Zwanghaftigkeit* zeigt sich z.B. in übermäßig risikovermeidendem Verhalten auf Basis einer Überschätzung möglicher zukünftiger Risiken, was zu einer Verzögerung von Entscheidungsfindungen führen kann, da zu lange überlegt wird. *Ambivalenz* wiederum zeigt sich für die Autoren in einer echten Reduktion der Anzahl an Entscheidungen einer Person, selbst bei nur zwei möglichen Alternativen. Sie beruht dabei auf evaluativen Konflikten, welche die Person nicht zu lösen imstande ist. Das Spektrum an Fehlformen umfasst an dem einen Ende demnach Personen, die Handlungsoptionen ohne Überlegung in Handlungen umsetzen (impulsiv oder voreilig Handelnde), an dem anderen Ende Personen, die viel zu lange über Handlungsoptionen nachdenken (zwanghaft oder übervorsichtig Handelnde), und schließlich noch solche, die selbst nach langem Nachdenken überhaupt keine Entscheidung treffen (ambivalente bzw. unentschlossene Personen).

Willensschwäche als Handeln wider das bessere Urteil fasse ich als der Phase 2 zugehörig an. Willensschwäche in diesem Sinne betrifft also nicht die Generierung von Handlungsoptionen oder die Ausführung oder Initiierung der ausgewählten Option. Genauer werde ich diese Art der

---

<sup>14</sup> Vgl. Kap. II.1.

<sup>15</sup> Vgl. hier Studien von George Ainslie (2004) über hyperbolische Diskontierung. Vgl. Kap. III.2.2.

Willensschwäche als *motivationale Willensschwäche* bezeichnen. Die philosophisch interessante Frage ist hier, wie eine Person, obwohl sie ein vergleichendes Urteil gefällt hat, dass eine bestimmte Handlungsoption besser sei als die anderen, trotzdem sich für die als weniger gut beurteilte Option entscheiden und so wider ihr besseres Urteil handeln kann. Von besonderer philosophischer Relevanz ist diese Frage dann, wenn die Bedingungen, unter denen dieses Phänomen auftreten sollte, idealisiert werden: Ist eine solche Willensschwäche, auch wenn die Überlegung nicht durch vorangegangene psychische Fehlleistungen wie die oben genannten beeinträchtigt wird, möglich? Kann jemand gewissermaßen sehenden Auges die schlechtere von zwei Handlungsoptionen wählen, wenn er oder sie nicht impulsiv, zwanghaft oder ambivalent ist?

*Phase drei* bildet die Initiierung von Handlungen, d.h. die beginnende Realisierung der gewählten Handlungsoption. In der philosophischen Literatur sind dies Fälle von sogenannter ‚last ditch akrasia‘ (Alfred Mele), ‚brazen akrasia‘ (David Pears) oder Handlungen gegen ein unconditionales evaluatives Urteil (Davidson). Aber auch einige Formen von Willensschwäche zählen darunter: Die Person hat in diesen Fällen Handlungsoptionen generiert und sich für eine entschieden und handelt dennoch nicht entsprechend. Ich bezeichne diese Form der Willensschwäche als *exekutive Willensschwäche*. Die Autoren unterscheiden prinzipiell zwei Arten dysfunktionaler Formen der Handlungsinitiierung: Fälle der beeinträchtigten Handlungsinitiierung in Gestalt von *Handlungsträgheit* (motor inertia) und Fälle der *Hyperaktivität und Handlungsimpulsivität* (motor impulsivity). Bei ersteren werden Handlungsinitiierungen unterlassen, bei letzteren unterbleiben Unterlassungen obwohl solche angebracht wären: die Person führt Handlungen aus, die sie dabei behindern, ihre Handlungen, für die sie sich entschieden hat, auszuführen. Im Unterschied zu Impulsivität in Phase 2 soll diese Art der Impulsivität die motorischen Fähigkeiten einer Person, Entscheidungen umzusetzen, betreffen, nicht die Art und Weise, wie sie Entscheidungen fällt.

Insgesamt ist dieses Schema der Klassifikation verschiedener Arten von Willensschwäche anhand der Lokalisation in der Stelle des Entscheidungsprozesses nützlich, um verschiedene Arten praktischer Irrationalität auseinanderzuhalten, die Diskussion pathologischer und alltäglicher Formen von praktischer Irrationalität in Psychologie und Philosophie einander anzunähern und einen einheitlichen begrifflichen Rahmen für eine solche zu liefern. Dabei ist zu bedenken, dass den verschiedenen Formen von Akrasia bzw. Willensschwäche in den unterschiedlichen Phasen nicht dieselben Eigenschaften zukommen, die in der Philosophie vorwiegend diskutiert werden – also freies, intentionales und irrationales Handeln wider das eigene bessere Urteil. Nur wenige der psychopathologischen Störungen sind dem reflexiven Zugriff seitens des Akteurs überhaupt zugänglich und betreffen Fehlfunktionen basaler psychischer und physischer Fähigkeiten. Ich werde mich in dieser Untersuchung von Willensschwäche vorwiegend auf Fälle beschränken, in denen die Person tatsächlich Überlegungen anstellt und über ein ansonsten normales Maß an

Handlungsoptionen (Phase 1) und normale Fähigkeiten der Handlungsinitiierung (Phase 3) verfügt. Zudem nehme ich an, dass die Überlegung und Abwägung der Handlungsoptionen zwar niemals frei ist von Beeinträchtigungen, wie z.B. in Form der Bevorzugung greifbarer kleiner Gewinne gegenüber zukünftig großen, dass diese Beeinträchtigung jedoch nicht automatisch dazu führt, dass es dem Akteur nicht möglich wäre, ein angemessenes vergleichendes Werturteil zu fällen. Der spezifische Beitrag der Philosophie besteht unter anderem darin, die Natur der Gründe, die in Phase 2 abgewogen werden, begrifflich zu fassen und zu verstehen, das Verhältnis von Evaluation und Motivation zu klären (dies betrifft natürlich alle drei Phasen), die Begriffe der Intentionalität und Freiheit von Handlungen zu bestimmen, sowie zu klären, in wie fern diese Verhaltens- und Handlungsweisen rational bzw. irrational sind. Darüber hinaus stellt sich natürlich die Frage, wie das Verhältnis der Person zu ihren Handlungen zu verstehen ist und was Selbstbestimmung in diesem Kontext bedeuten soll.

Gegen diese Art der Herangehensweise an diese Fragen kann mehreres eingewendet werden. *Erstens* kann generell der individualistische Zugriff auf das Phänomen des Handelns und seiner irrationalen Fehlformen kritisiert werden. Personen handeln nur selten alleine, und auch wenn sie dies tun, so sind ihre Handlungen, ja ihre gesamte Handlungsfähigkeit geprägt durch interaktive Muster und den Ergebnissen früherer Interaktion. Niemand würde bestreiten, dass wir alle in gewissem Sinne erst lernen müssen, zu handeln, und dass die meisten Handlungen kooperativ und im Zusammenspiel mit anderen Personen geschehen. Doch auch wenn dies richtig ist, so ist die Klärung dieser Fragen, die Willensschwäche aufwirft, aus der Perspektive des Akteurs legitim. Wir haben ein Interesse daran, die Perspektive des Einzelnen zu verstehen und die Art und Weise, wie der Einzelne handelt und überlegt. Dies ist ein legitimes Erkenntnisinteresse und ich denke, dass für die Erforschung der damit zusammenhängenden Fragen die Perspektive des Akteurs zentral ist.

*Zweitens* kann eingewendet werden, dass diese Herangehensweise Handlungen und Akteure aus ihrem umgreifenden Kontext löst und damit das gesamte Phänomen auf bloß abstraktem Niveau behandelt. Natürlich tun Personen in irgendeiner Weise immer gerade etwas, während sie Überlegungen über zukünftige Handlungen anstellen. Personen haben nicht nur einzelne Handlungsabsichten in konkreten Situationen, sondern übergreifende kleinere und größere Pläne, in denen einzelne Handlungen nur Episoden darstellen. Diese größeren und kleineren Pläne haben natürlich Einfluss auf die Interpretation der jeweiligen Handlungssituation sowie auf die Überlegungen darüber, was zu tun ist. Zurecht hat unter anderem Michael Bratman auf diese Dimension von Intentionen als mehr oder weniger konkrete Pläne hingewiesen und deren Rolle

für praktische Überlegungen dargelegt.<sup>16</sup> Frühere Handlungen und Pläne restringieren bereits *extern* auf eine fundamentale Weise, über welche Handlungsoptionen eine Person überhaupt nachdenkt, sie prägen demnach die Phase der Generierung von Handlungsoptionen. Wo dies nicht geschieht, da kann eine Art der oben dargestellten Beeinträchtigung des Entscheidungs- und Handlungsprozesses vorliegen: eine Person agiert dann ziel- oder richtungslos in einer Situation, hat gewissermaßen ihre Orientierung verloren, weil sie keinerlei oder viel zu viele Handlungsoptionen generiert, da sie keine festen Pläne hat, die ihr Leben einen Rahmen geben können. Außerdem restringieren Pläne *intern* die Überlegungen hinsichtlich der Abwägung verschiedener so generierter Handlungsoptionen, indem sie Unverträglichkeiten verschiedener Handlungsoptionen mit übergreifenden Plänen erzeugen. Wenn ich den Plan habe, nächste Woche auf Urlaub zu fliegen, so kann ich nicht zusagen, auf die Geburtstagsparty eines Freundes zu kommen, die nächste Woche stattfinden soll. Aber dennoch spricht dies nicht prinzipiell dagegen, auch die Struktur einzelner Überlegungen und die Eigenart einzelner Abwägungen verschiedener Handlungsoptionen zu betrachten, solange der umfassende Kontext nicht aus dem Blickfeld gerät und man sich der involvierten Abstraktion bewusst bleibt.

*Drittens* kann eingewendet werden, dass Willensschwäche eine dispositionale Eigenschaft von Personen sei, und deshalb überhaupt nicht auf Basis einer einzelnen Handlung vorliegen könne. Dem stimme ich prinzipiell zu. Willensschwäche ist eine Eigenschaft von Personen, die sich über mehrere Handlungen zeigt, und als solche kaum auf Basis einer einzelnen Handlung zuschreibbar ist. Dennoch können wir diese Disposition nur verstehen, wenn wir sie auch in den einzelnen Fällen, in denen Personen gegen ihr besseres Urteil handeln, verstehen, denn irgendwann muss eine Person willensschwach gehandelt haben um zurecht als willensschwach zu gelten. Auch ist es so, dass in den Fällen von Willensschwäche als klarsichtiges Handeln wider das bessere Urteil, die Person charakteristischerweise sich selbst Willensschwäche zuschreibt, ihre Willensschwäche bemerkt und Gefühle der Reue pflegt. Es ist denkbar, dass die Person dies aufgrund einer einzelnen Handlung, die Folge einer wichtigen Entscheidung ihres Lebens war, tut. Also auch wenn Willensschwäche ein dispositionale Eigenschaft von Personen ist, so müssen wir verstehen, was es heißt, dass diese Person diese Eigenschaft zeigt, und dies führt zu der Frage, was in einzelnen willensschwachen Handlungen genau geschieht und wie diese überhaupt möglich sind.

### **I.3 Akrasia oder Willensschwäche?**

Nach diesen kurzen methodologischen Ausführungen möchte ich zurückkommen auf die Frage der Bestimmung des Phänomens der Willensschwäche. In der philosophischen Diskussion wurde

---

<sup>16</sup> Vgl. Bratman 1987.

vor allem die Frage gestellt, ob nicht Akrasia von Willensschwäche streng zu scheiden sei, da sie verschiedene Phänomene betreffen. So schreibt etwa Richard Holton:

„Whenever I have asked nonphilosophers what they take weakness of will to consist in, they have made no mention of judgments about the better or worse course of action. Rather, they have said things like this: weak-willed people are irresolute; they do not persist in their intentions; they are too easily deflected from the path that they have chosen. My aim here is to pursue this line of thought. I shall develop the idea that the central cases of weakness of will are best characterized not as cases in which people act against their better judgment, but as cases in which they fail to act on their intentions.” (Holton 1999, 241)

Holton plädiert also dafür, Willensschwäche im Unterschied zu Akrasia nicht als Handeln wider das bessere Urteil, sondern als vorschnelle Revision bestehender Handlungsintentionen aufzufassen. Willensschwäche betreffe dem vorherigen Schema nach Phase 3, während Akrasia auf Phase 2 beschränkt bliebe. Dagegen hat Alfred Mele eingewendet, dass empirische Studien der Alltagssprache keine eindeutige Unterscheidung zwischen Willensschwäche als vorschnelles Revidieren vorhandener Intentionen und Akrasia als Handeln wider besseres Urteil zulasse.<sup>17</sup> Ich denke, dass diese beiden Arten praktischer Irrationalität auseinandergehalten werden sollten, ohne mich darauf festzulegen, welches der Phänomene empirisch betrachtet öfters auftritt als das andere. Philosophisch gesehen ist das Phänomen der motivationalen Willensschwäche interessanter, da es wichtige Fragen über den Zusammenhang zentraler Begriffe aufwirft.<sup>18</sup> Die schließt aber eine philosophische Beschäftigung mit exekutiver Willensschwäche natürlich nicht aus.

Es bleibt die Frage, ob nicht alle Fälle von exekutiver Willensschwäche auf Fälle von motivationaler Willensschwäche reduzierbar sind. Ob die Revision einer Entscheidung als Handlung wider das bessere Urteil aufzufassen ist, entscheidet der Umstand, ob wirklich ein Urteil des Akteurs über die Revision seiner Intentionen vorliegt oder nicht. Liegt ein solches vor, dann wäre dies ein Fall von motivationaler Willensschwäche. Liegt kein explizites Urteil vor, so sollten wir diese Fälle als andere Form praktischer Irrationalität behandeln. Diese Diskussion ähnelt der Diskussion um die Adäquatheit der Klassifikation Rortys, da auch sie nicht sauber zwischen Fällen trennt, in denen der Akteur ein explizites Urteil in ihren Überlegungen fällt, und solchen, in denen dieses Urteil nur dispositional zugeschrieben werden kann. Ungeachtet dessen sollten wir Fälle, in denen jemand gegen seine in Überlegungen getroffenen Entscheidungen (Absichten) handelt, anders analysieren als solche, in denen er angesichts dieser Überlegungen und tatsächlich vorliegenden vergleichenden Werturteilen über die Gründe für bestimmte Handlungsoptionen eine andere Entscheidung trifft als die durch dieses Urteil präferierte. Alfred Mele hingegen zielt auf einen weiten Begriff von Akrasia als fehlender Selbstkontrolle ab, der alle

---

<sup>17</sup> Vgl. Mele 2010a.

<sup>18</sup> Dies tut auch Mele in seiner Unterscheidung zwischen evaluativen und exekutiven praktischen Selbstverpflichtungen (evaluative und executive practical commitments). Vgl. Mele 2010a, 394.

möglichen Arten dieser Fehlleistung umfasst. Aufgrund der Diskussion des Schemas von Kalis & al. meine ich, dass die jeweils in Frage stehende Art der Selbstkontrolle von Phase zu Phase sehr verschieden ist und nicht allesamt nur in der Struktur der Willensschwäche als Intentionenbildung entgegen des vorliegenden besseren Urteils zu verstehen sind. Wo ein solches Urteil nicht vorliegt, sind andere Fähigkeiten des Akteurs zur Selbstkontrolle betroffen, die unterschiedliche kognitive, volitionale und motorische Leistungen umfassen. Jemand kann auf sehr viele verschiedene Arten gegen sein besseres Urteil handeln, und diese sollten in der Analyse voneinander getrennt werden. Dies bedeutet natürlich nicht, dass Fälle von Willensschwäche, wie sie z.B. Holton beschreibt, philosophisch uninteressant wären. Was ihnen jedoch im Unterschied zu Fällen von Willensschwäche, wie ich sie hier vorwiegend verstehe, mangelt, ist die charakteristische reflexive Dimension der letzteren. Wo diese vorliegt, sollten wir sie nicht als intentionales Handeln gegen eigene Entscheidungen (im Sinne von Absichten) oder als Unfähigkeit, Intentionen in Handlungen umzusetzen, analysieren.

#### **I.4 Willensschwache Überzeugungen**

Eine weitere wichtige Diskussion umfasst die Ausdehnung des Phänomens der Willensschwäche auf Bereiche theoretischer Irrationalität. Dies betrifft vor allem die Frage, ob willensschwache Überzeugungen (*akratic belief*) möglich sind.<sup>19</sup> Bereits bei Rorty haben wir gesehen, dass sie auch bei klassischerweise theoretischen kognitiven Leistungen wie Wahrnehmungen und Überzeugungen das Auftreten von Willensschwäche annimmt, und zwar dann, wenn die in Frage stehenden theoretischen Einstellungen Produkte von absichtlichem Handeln sind und der Akteur ein präferiertes Urteil hat, die gegen diese Handlungen sprechen. So könnte eine Person das Urteil haben, dass sie immer erst alle relevanten Evidenzen, die für oder gegen eine Überzeugung sprechen, abwägt, und auch keine dieser Evidenzen unterdrückt, um ein möglichst ausgewogenes Urteil zu fällen. Dennoch können Personen, die diese Auffassung teilen, Überzeugungen als Folge einseitiger Suche nach Evidenzen, die für diese Überzeugung sprechen, haben. In diesen Fällen, so Rorty, läge Willensschwäche vor, da die Person fähig wäre, die fraglichen Handlungen, die zu dieser Überzeugung führen, zu unterlassen bzw. zu ändern, und zudem das präferierte Urteil hat, dass gegen eine solche Art, Überzeugungen zu bilden, spricht. Auch hier wiederum denke ich, dass es sinnvoll ist, zwischen den Fällen zu unterscheiden, in denen jemand tatsächlich ein Urteil in einer Überlegung trifft, dass es z.B. besser wäre, nach bestimmten Evidenzen zu suchen, und sich dann dennoch dafür entscheidet, nicht nach ihnen zu suchen, von Fällen, in denen ein solches Urteil nicht vorliegt. Der erste Fall würde in den hier behandelten Bereich von Willensschwäche fallen, der zweite nicht.

---

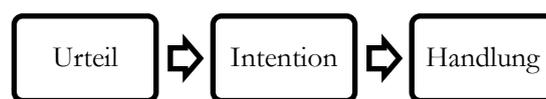
<sup>19</sup> Vgl. David Owens 2002; Ribeiro 2011;

Eine andere Frage ist, ob eine Person eine Überzeugung nicht nur indirekt, als Folge gewisser Handlungen, sondern direkt auf willensschwache Weise bilden könnte. Hier gibt es jedoch starke Einwände, die gegen eine solche Annahme sprechen. Personen können sich nicht in einem starken Sinne dafür entscheiden, etwas Bestimmtes zu glauben. Sie können gewissen Gründen für ihre Überzeugungen mehr Beachtung schenken, sie können nach Inkohärenzen und Widersprüchen in und zwischen ihren Überzeugungen suchen, aber sie können nicht einfachhin etwas glauben oder nicht glauben, bloß weil sie das wollen.<sup>20</sup> Fälle von Wunschdenken und Selbsttäuschung verlangen offenbar eine andere Analyse als Willensschwäche und Fälle irrationalen Handelns, auch wenn sie in vieler Hinsicht sehr ähnlich sein mögen.

## I.5 Der Begriff der Willensschwäche

### I.5.1 Definition

Die vorhergehende Diskussion diene dazu, mit dem Phänomen der Willensschwäche vertraut zu machen, wie es sich innerhalb verschiedener theoretischer und praktischer Zugriffe zeigt. Willensschwäche ist ein vielschichtiges Phänomen und der zugehörige Begriff ist daher *komplex*. Dennoch ist es ratsam und möglich, verschiedene Komponenten des Begriffs bzw. Aspekte des Phänomens der Willensschwäche zu isolieren und begrifflich zu fixieren, um ein adäquateres Verständnis von Willensschwäche zu ermöglichen. Für diese Fixierung ist es erforderlich, zumindest zwei grundlegende Typen von Willensschwäche voneinander zu unterscheiden, abhängig von der Stelle im Prozess vom Urteil eines Akteurs über seine Handlungsoptionen zur Ausführung einer Handlung (Urteil – Intention – Handlung): motivationale und exekutive Willensschwäche.



*Motivationale Willensschwäche:* Eine Person urteilt, dass sie eine Handlung tun sollte, oder dass eine ihr offenstehende Handlung besser wäre als eine andere, und intendiert dennoch eine davon verschiedene, alternative Handlung.

Eine Person P zeigt *motivationale* Willensschwäche dann und nur dann, wenn

- 1) P *glaubt*, dass sie selbst F-en sollte oder dass F-en besser wäre als G-en, und dass es ihr offensteht zu F-en oder zu G-en (Evaluation)
- 2) P *intendiert* zu G-en (Intentionalität)

---

<sup>20</sup> Dagegen ist der doxastische Voluntarismus eine Theorie, die von einem starken Anteil des Willens an der Bildung von Überzeugungen ausgeht. Für eine Kritik an diesem vgl. Williams 1970.

- 3) P *verantwortlich* ist für die Intention zu G-en (Verantwortung)
- 4) P *prima facie kritisierbar* ist für ihre Intention zu G-en (Normativität)

*Exekutive Willensschwäche*: Eine Person intendiert eine Handlung, und vollzieht eine andere oder vollzieht die intendierte Handlung nur halbherzig oder unzureichend.

Eine Person P zeigt *exekutive* Willensschwäche dann und nur dann, wenn

- 1) P *intendiert* zu F-en (Intention)
- 2) P *intentional* G-t oder nur halbherzig und unzureichend F-t (Handlung)
- 3) P *verantwortlich* ist für ihr G-en/F-en (Verantwortung)
- 4) P *prima facie kritisierbar* ist für ihr F-en/G-en (Normativität)

Diese vier Bedingungen der beiden Formen tauchen in unterschiedlicher Gewichtung in den verschiedenen Definitionen von Willensschwäche immer wieder auf. Obwohl die einzelnen Bedingungen oder Begriffsteile natürlich in engen Beziehungen zueinander stehen, ist es dennoch ratsam, sie analytisch voneinander zu trennen und die Probleme, die mit diesen Begriffselementen speziell verbunden sind, einzeln zu diskutieren, um ein adäquateres Verständnis von Willensschwäche zu gewährleisten. Diese Definition ist weit genug, um genügend Platz für die verschiedenen Formen von Willensschwäche zu bieten, die zuvor besprochen wurden. Sie ist aber auch eng genug, indem sie die Unterscheidung von Willensschwäche zu anderen, benachbarten Phänomenen ermöglicht.

*Erstens* lässt die Formulierung ‚x zeigt Willensschwäche‘ offen, ob Willensschwäche als Charakterdisposition oder als Handlungstypus verstanden wird. Willensschwäche als Charakterzug zeigt sich in einzelnen willensschwachen Handlungen, und einzelne willensschwache Handlungen formen die Disposition der Willensschwäche. Allgemein sind Zuschreibungen von Charakterdispositionen epistemologisch wohl nur durch den Hinweis einzelner Aktualisierungen zu rechtfertigen; umgekehrt leistet die Zuschreibung einer Disposition mehr als die bloße Summierung einzelner Aktualisierungen, da sie explanatorische, normative und praktische Funktionen erfüllt.

*Zweitens*: Die Wissens- bzw. Glaubensbedingung muss in einem weiten Sinne von Wissen bzw. Glauben verstanden werden. Sowohl der Grad der Überzeugung bzw. der Meinung als auch ihr eher dispositionaler oder aktueller Charakter sind variabel. Die praktische Signifikanz von Willensschwäche hängt dabei vom Grad der Überzeugung, dessen Inhalt, die mit dem Inhalt verbundene normative Signifikanz sowie der Grad der Habitualisierung der Überzeugung ab. Philosophisch am stärksten umstritten sind Fälle von Willensschwäche, in denen die

Überzeugung der Person aktuell gegeben, sie stark geglaubt wird und inhaltlich Handlungen betreffen, die normativ wichtig sind für die Person und ihr soziales Umfeld. Die Frage, ob Willensschwäche im Handeln wider besseren *Glauben* oder im Handeln wider besseres *Wissen* bestehe, bleibt an dieser Stelle noch offen. Diesem Umstand wird die schwache Formulierung ‚X glaubt‘ gerecht.

*Drittens* ist die willensschwache Handlung intentional. Für Theorien, die Handlungen notwendig als intentional auffassen, ist diese Bedingung bereits erfüllt, wenn etwas als ‚Handlung‘ charakterisiert wird; der Zusatz ‚intentional‘ ist für diese Theorien redundant. Dennoch schadet es nicht, diesen redundanten Bestandteil explizit zu machen, noch dazu, wenn verschiedene Auffassungen über die Intentionalität von Handlungen bestehen. Diese Bedingung impliziert nicht, dass die Person intentional oder absichtlich willensschwach wäre; sie handelt lediglich so, dass sie durch ihr Handeln Willensschwäche zeigt, bzw. drücken ihre intentionalen Handlungen ihren Charakterzug der Willensschwäche aus. Was genau mit ‚intentional‘ gemeint ist, bleibt auch hier vorerst offen. Klar davon unterschieden sind jedoch Verhaltensweisen, die unter keiner Beschreibung als intentional beschrieben werden können, wie z.B. reine Körperbewegungen hervorgerufen durch äußere Kräfte oder rein körperliche Reflexe. Intentionalität oder Absichtlichkeit sollte auch hier vorerst eher schwach verstanden werden und nicht im Sinne von vorsätzlich oder überlegt. Nicht jede intentionale Handlung ist notwendigerweise vorsätzlich und umgekehrt ist auch nicht jede vorsätzliche Handlung unter der Beschreibung intentional, unter der sie intendiert wurde. Welche epistemischen Zustände die Person haben muss, damit sie intentional handeln kann, bleibt auch offen. Viele meinen, dass eine Person zumindest glauben muss, dass eine Handlung möglich ist, um sie auch intendieren zu können. Umgekehrt würden also Handlungen, von denen die Person selbst glaubt, dass sie diese unmöglich ausführen könnte, keinen willensschwachen Konflikt hervorrufen, da eine Person solche Handlungen nicht intendieren könnte.

*Viertens* trägt die willensschwache Person Verantwortung für ihre Willensschwäche. Dies kann verstanden werden als Verantwortung für ihre willensschwachen Handlungen als auch für ihre Willensschwäche als Charakterzug. Die Verantwortung für Willensschwäche ist nicht von vornherein gleichzusetzen mit der Intentionalität der willensschwachen Handlung. Dies ist vielmehr eine substantielle These über die Verantwortlichkeitsbeziehung und muss eingehend diskutiert werden.

*Fünftens* ist die normative Dimension notwendiger Bestandteil von Willensschwäche. Willensschwäche ist kein rein deskriptives Konzept, sondern wesentlich normativ. Wir kritisieren Personen für ihre willensschwachen Handlungen und für ihre Willensschwäche, was zu der Frage

führt, ob und, wenn ja, aus welchen Gründen diese Kritik gerechtfertigt werden kann. Die Qualifizierung durch ‚prima facie‘ ist deswegen notwendig, da es willensschwache Handlungen geben kann, die normativ geboten sind, und in diesen Fällen wäre es verfehlt, die Person schlechthin für die Ausführung der Handlung zu kritisieren. So könnte jemand meinen, dass es alles in allem besser wäre, einen Bankraub auszuführen, und diesen Plan aus Willensschwäche nicht fassen oder bei der Ausführung kalte Füße bekommt. Hinsichtlich der Willensschwäche ist die Person kritisierbar, aber schlechthin liegt es an den anderen gegebenen Gründen, die Person für die Ausführung der Handlung zu kritisieren. In einem minimalen Sinne kritisierbar sind willensschwache Handlungen als subjektiv irrational, als eigene Standards des rationalen Handelns widersprechend. Natürlich besteht hier die Schwierigkeit, wie genau diese Standards zu verstehen sind. Substanziell wird die Kritik dann, wenn sie mit Überlegungen über das gute Leben und der Rolle von rationaler Selbstbestimmung für ein solches Leben verknüpft wird. Auch diese Frage bleibt in dieser vorläufigen Bestimmung offen für die Diskussion.

### **1.5.2 Dimensionen von Willensschwäche**

Anhand der vorhergehenden Diskussion lassen sich verschiedene Dimensionen von Willensschwäche unterscheiden, welche für die philosophische Diskussion von Relevanz sind und das Phänomen begrifflich weiter differenzieren. Dabei lassen sich Differenzierungen auf der Gegenstandsseite (a-d) von solchen auf Seiten des methodischen Zugangs (e-g) unterscheiden.<sup>21</sup>

#### ***a) Zeitliche Dimension: synchron/diachron***

Mit synchroner Willensschwäche ist gemeint, dass die Person wider ihren besseren Glauben handelt, der ihr aktuell gegeben oder dispositional zugeschrieben werden kann. Dies wird auch oft als stärkste Form der Willensschwäche bezeichnet, da sie von vielen als die philosophisch problematischste Form betrachtet wird. Bei diachroner Willensschwäche handelt die Person im Einklang mit einem vorliegenden dispositional zuschreibbaren oder aktuell gegebenen Glauben, der jedoch im Widerspruch zu einem früheren Glauben steht. Die Person revidiert demnach kurzfristig ihr Urteil über die bestehenden Handlungsalternativen und handelt diesem neuen Urteil entsprechend. Diese Form von Willensschwäche kann auch als kurzfristige Urteilsrevision oder als kurzfristige (irrationale) Änderung von Präferenzen bezeichnet werden.

#### ***b) Träger von Willensschwäche: episodisch/dispositional***

Die Willensschwäche kann als Typus von Handlungen verstanden werden oder als Disposition, z.B. als Charakterzug, bestimmte Handlungen auszuführen. Im ersten Fall ist der Träger der Eigenschaft der Willensschwäche eine einzelne Handlung, im zweiten die handelnde Person.

---

<sup>21</sup> Vgl. Müller 2009, 18ff.

### **c) Art der Willensschwäche: motivational/exekutiv**

Mit motivationaler Willensschwäche wird der Fall des Handelns wider ein besseres Urteil verstanden. Dagegen meint exekutive Willensschwäche das Handeln wider eine bestehende Intention bzw. das Unterlassen der Ausführung einer Intention, etwas zu tun. Die erste Form betrifft demnach die Fähigkeit, Intentionen auf Basis von Urteilen zu fällen, die zweite die Fähigkeit, solche Intentionen auch wirklich in die Tat umzusetzen.

### **d) Inhalt des Urteils: moralisch/prudentiell**

Diese Unterscheidung betrifft den Inhalt des Glaubens, gegen den die willensschwache Person handelt. Von moralischer Willensschwäche kann gesprochen werden, wenn das entsprechende Urteil bzw. der Glaube moralische Normen oder Vorschriften betrifft. Davon unterschieden wäre prudentielle Willensschwäche als Handeln wider Handlungsurteile auf Basis von Klugheitsgründen. Diese Unterscheidung macht natürlich nur Sinn vor dem Hintergrund, dass sich moralische Gründe überhaupt von prudentiellen Gründen abgrenzen lassen, mithin die Moral (oder die Klugheit) einen eigenständigen Bereich bildet.

### **e) kognitiv/konativ**

Hier handelt es sich um eine Unterscheidung zwischen verschiedenen Theorietypen von Willensschwäche. Rein kognitive Theorien verweisen in ihren Explikationen und Erklärungen von Willensschwäche nur auf kognitive Elemente, d.h. Urteilsgehalte und Glaubensakte. Rein konative Theorien verweisen demgegenüber nur auf konative Momente wie Wünsche und Strebungen, um Willensschwäche zu explizieren und zu erklären. Gemischte Theorien verweisen sowohl auf kognitive als auch auf konative Aspekte, wobei das konkrete Verhältnis zwischen den beiden Elementen diese Theorien wieder voneinander unterscheiden. Die hier vorgeschlagene Theorie ist eine kognitiv-konative.

### **f) explikativ/explanatorisch**

Eine explikative Theorie setzt sich nur zum Ziel, den Begriff des willensschwachen Handelns bzw. der Willensschwäche auseinander zu legen. Explanatorische Theorien bieten Erklärungen für das Auftreten von willensschwachen Handlungen bzw. Willensschwäche an. Philosophische Theorien, die keine empirische Forschung umfassen, können nach dem hier vorliegenden Verständnis solchen explanatorischen Theorien nur durch explikative Klärung zuarbeiten, erheben aber selbst keine Erklärungsansprüche abseits der Erläuterung von ganz allgemeinen explanatorischen Strukturen von Willensschwäche.

### **g) deskriptiv/normativ/evaluativ**

Eine rein deskriptive Theorie der Willensschwäche beschreibt lediglich das Auftreten willensschwacher Handlungen und/oder die Disposition, solche Handlungen auszuführen. Eine normative Theorie der Willensschwäche untersucht, weshalb Personen willensschwaches Handeln oder Willensschwäche vermeiden sollen, indem sie zeigt, welche Normen gegen ein solches Handeln sprechen. Es ist strittig, ob es eine rein deskriptive Theorie überhaupt geben kann. Dies hängt hauptsächlich davon ab, ob man den zugrundeliegenden Begriff der Rationalität und des Handelns nach Gründen deskriptiv oder normativ versteht. Die evaluative Dimension von Willensschwäche betrifft die Frage nach den spezifischen Werten, die mit Willensschwäche in Verbindung stehen. Welche Werte werden durch willensschwaches Handeln realisiert und welche werden nicht realisiert? Die hier vorgeschlagene Theorie sieht Selbstrespekt bzw. Selbstachtung als zentralen Wert an, der durch willensschwaches Handeln und die Tendenz zur Willensschwäche untergraben wird. Da Selbstrespekt ein fundamentaler (vielleicht der fundamentalste) Wert ist, der zudem nahtlos mit unserem Verständnis von uns selbst als Personen verbunden ist, berührt Willensschwäche eine grundlegende Dimension unserer Art zu leben. Die normative Frage, weshalb jemand nicht willensschwach handeln sollte, werde ich zweifach beantworten: in einem schwachen Sinn ist es unmöglich, eine Person zu sein, die *immer* willensschwach handelt. Es macht demnach keinen rechten Sinn aus der Perspektive einer Person zu fragen, weshalb sie überhaupt Willensschwäche vermeiden sollte. Diese Antwort ist jedoch sehr schwach und erlaubt, dass Personen nicht nur hin und wieder, sondern sogar bis zu einem sehr hohen Grad ein willensschwaches Leben führen. Weshalb sie dies vermeiden sollten, kann nur durch ein substanzielles Verständnis eines idealen Lebens beantwortet werden, im Rückgriff auf ein Verständnis von Selbstrespekt und des damit verbundenen Ideals, sein Leben zu *führen* anstatt es nur *vor sich hinzuleben*. Diese substanzielle Konzeption mag nicht alle überzeugen, und der normative Anspruch bleibt daher relativ gering. Doch dies stimmt mit unseren alltäglichen Intuitionen gut überein: Wir zwingen niemanden letztlich, nicht willensschwach zu sein. Wir mögen es bedauern, die willensschwache Person verurteilen, aber sehen es nicht als gerechtfertigt an, die Person gegen ihren Willen zu disziplinieren bzw. sie Maßnahmen zu unterziehen, mit dem Ziel, ihre Willensstärke zu erhöhen, wenn sie nicht in diese Maßnahmen einwilligt oder uns zu diesen ermuntert bzw. auffordert.

### **I.5.3 Abgrenzungen**

Das hier vorgeschlagene Konzept von Willensschwäche versucht eine gemeinsame minimale Basis für die Debatte um das rechte Verständnis des Phänomens der Willensschwäche zu bieten. Jede Unterscheidung muss natürlich auch hinsichtlich des davon Unterschiedenen betrachtet und gerechtfertigt werden. Das Phänomen der Willensschwäche kann durch die vier Bedingungen

(Evaluation, Intentionalität, Verantwortung, Normativität) positiv charakterisiert werden. Negativ lässt sich Willensschwäche von anderen benachbarten Phänomenen abgrenzen: Selbsttäuschung, Versuchung, Depression, Sucht und Zwangsstörung.

### **a) Selbsttäuschung**

Die philosophische Debatte um Selbsttäuschung ist mindestens ebenso alt und umfangreich wie die Debatte um Willensschwäche. Beide Phänomene teilen miteinander ihren prekären Charakter und den Anschein der Unvernünftigkeit. Prekär deshalb, da auch der Begriff der Selbsttäuschung unter dem Verdacht steht, leer zu sein, wenn damit das Phänomen gemeint wird, dass eine Person zugleich etwas wisse und dieses Wissen absichtlich vor sich selbst verberge. Unvernünftig deshalb, da ein solcher Zustand dem rationalen Ziel jedes Erkenntnistrebens zuwiderlaufe. Während Willensschwäche jedoch in Handlungen wider besseres Wissen besteht, so Selbsttäuschung im Glauben wider bessere Evidenz. Das Verhältnis zwischen Selbsttäuschung und Willensschwäche kann grundlegend auf drei verschiedene Weisen verstanden werden:

- 1) Willensschwäche und Selbsttäuschung sind voneinander distinkte Phänomene und keines kann auf das andere reduziert werden.
- 2) Willensschwäche ist in Wahrheit eine Form der Selbsttäuschung und kann daher analytisch auf diese reduziert werden. Die Person täuscht sich darüber, was sie tun sollte oder was sie am meisten tun will.
- 3) Selbsttäuschung ist in Wahrheit eine Form der Willensschwäche und kann daher analytisch auf diese reduziert werden. Die Person ist willensschwach hinsichtlich ihres Wahrheitswillens.

Die genauere Diskussion von Theorien des Typs 2 erfolgt später.<sup>22</sup> Da es hier nicht um eine Analyse der Selbsttäuschung geht, kann die Diskussion von Theorien des Typs 1 und 3 hier unterbleiben, außer wir lernen dabei etwas über das Phänomen der Willensschwäche, nicht das der Selbsttäuschung.

### **b) Versuchung und Verführung**

Das Phänomen der Versuchung wie der Verführung kann aus der Perspektive des Versuchers bzw. Verführers oder der Person, die in Versuchung geführt bzw. verführt wird, verstanden werden. Da jedoch nicht jeder Fall von Versuchtsein bzw. Verführtwerden<sup>23</sup> notwendig eine andere Person involviert, sollte das Phänomen primär aus der Perspektive des oder der

---

<sup>22</sup> Vgl. Kap. III.1.

<sup>23</sup> Ich halte die beiden Begriffe ‚Versuchung‘ und ‚Verführung‘ für im Grunde synonym und werde sie in Folge so verwenden. ‚Verführung‘ wird vielleicht nur in interpersonalen Fällen gebraucht – ‚sich selbst verführen‘ klingt nicht richtig zumindest für meine Ohren, im Gegensatz zu ‚sich selbst versuchen bzw. in Versuchung führen‘ – und ist stärker sexuell konnotiert.

Versuchten bzw. Verführten verstanden werden. Außerdem können die in der Verführung geweckten oder angestachelten Begierden und Wünsche trivialer Natur oder moralisch relevant sein.<sup>24</sup> Wenn das Verführtwerden darin besteht etwas zu tun, von dem man zumindest glaubt, dass es in einem weiten Sinne nicht richtig ist zu tun, dann fällt das Phänomen mit synchroner motivationaler Willensschwäche zusammen, zumindest auf Seiten des Willensschwachen. Wenn Verführtwerden jedoch primär darin besteht, sein Urteil darüber, was man am besten tun sollte, zu ändern, dann ist es ein Fall von diachroner motivationaler Willensschwäche.<sup>25</sup> Die Verführung durch eine andere Person kann analog als Ausnutzung der Willensschwäche eines anderen verstanden werden. Spezielle philosophische Probleme stellen sich demnach für das Phänomen der Verführung nur in ihrer interpersonalen Gestalt als moralphilosophische Probleme des rechten Umgangs mit anderen Personen – dabei geht es primär um die Frage, ob Verführung unmoralisch ist und wenn ja, weshalb.

### **c) Depression**

Laut ICD-10-GM<sup>26</sup> (2013) umfasst die Gruppe der affektiven Störungen solche „Störungen deren Hauptsymptome in einer Veränderung der Stimmung oder der Affektivität entweder zur Depression – mit oder ohne begleitende(r) Angst – oder zur gehobenen Stimmung bestehen.“ (ICD 10 GM 2013, F30) Die von diesen Störungen betroffenen Personen leiden zudem an einem allgemeinen Schwund (Depression) oder Anstieg (Manie) des Aktivitätsniveaus ausgelöst durch die Stimmungsveränderung. Die Zuschreibung von Willensschwäche setzt im Unterschied zu diesen Phänomenen ein normales Affektivitäts- und Aktivitätsniveau voraus. Die genaue Abgrenzung wird letztlich graduell sein, auch bestimmt über das als gesellschaftlich *normal* akzeptierte Affektivitätsniveau.

### **d) Sucht**

Dieser Gruppe psychischer Störungen und Verhaltensstörungen durch psychotrope Substanzen ist gemein, dass sie in Verbindung zu einer spezifischen psychotropen Substanz stehen. Darunter zählen Phänomene wie Missbrauch, Abhängigkeitssyndrom, Entzug usw. mit jeweils für die entsprechenden Substanzen typischen Verlaufs- und Intoxikationsformen. Willensschwäche spielt innerhalb dieser Phänomene sicherlich eine gewisse Rolle. Dennoch gilt auch hier, dass die

---

<sup>24</sup> „One sense of temptation is used casually to refer to garden variety desires and attractions, and another, the main moral sense, connotes a desire for what one thinks is in some way wrong. I have claimed, furthermore, that analyzing temptation from the standpoint of a tempter, and taking deliberate interpersonal temptations as paradigmatic, ultimately misconstrues temptation by under-appreciating the essential role of the subject of temptation. Without an adequate phenomenology of what it is to *be* tempted, temptation as a relation between a tempter and a person tempted cannot be understood.“ (Hughes 2002, 104)

<sup>25</sup> Vgl. Holton 2009, ch. 5.

<sup>26</sup> Die deutsche Version der International Classification of Diseases, der international anerkannten Standardklassifikation für Krankheiten von der WHO (World Health Organisation) der UNO.

Zuschreibung von Willensschwäche vor dem Hintergrund eines als normal angesehenen Affektivitäts- und Aktivitätslevels geschieht. Dies heißt nicht, dass Süchtige nicht auch willensschwach sein können, aber die Zuschreibung des Vorliegens einer Sucht verändert die normative Bewertung aufgrund der Änderung der Verantwortlichkeitszuschreibung zu einzelnen Handlungen: Süchtige sind graduell in ihrer Freiheit eingeschränkt, ihrem besten Urteil gemäß zu handeln, im Unterschied zu Willensschwachen, bei denen eine solche Einschränkung nicht unbedingt vorliegen muss.

### **e) Zwangsstörung**

Kennzeichen von Zwangsstörungen und anderen zwanghaften Verhaltensstörungen sind immer wiederkehrende Gedanken und Handlungen, die zwar als einem selbst zugehörig, aber nicht als sinnvoll erlebt werden.

„Wesentliche Kennzeichen sind wiederkehrende Zwangsgedanken und Zwangshandlungen. Zwangsgedanken sind Ideen, Vorstellungen oder Impulse, die den Patienten immer wieder stereotyp beschäftigen. Sie sind fast immer quälend, der Patient versucht häufig erfolglos, Widerstand zu leisten. Die Gedanken werden als zur eigenen Person gehörig erlebt, selbst wenn sie als unwillkürlich und häufig abstoßend empfunden werden. Zwangshandlungen oder -rituale sind Stereotypen, die ständig wiederholt werden. Sie werden weder als angenehm empfunden, noch dienen sie dazu, an sich nützliche Aufgaben zu erfüllen.“ (ICD 10 GM 2013, F42)

Im Unterschied zu willensschwachen Handlungen ist das Moment des Zwanges, dem sich die Person ausgesetzt fühlt, wesentlich. Die Person fühlt sich selbst nicht frei, die jeweiligen Zwangshandlungen *nicht* auszuführen bzw. die jeweiligen Zwangsgedanken *nicht* zu denken. Letztlich wird die genaue Abgrenzung auch hier graduell sein und die jeweilige Dynamik der Entwicklung einer Person berücksichtigen müssen. So mag es sein, dass das Verhalten zu Anfang lediglich noch als willensschwach beschrieben und erlebt wird, bevor der Charakter des Zwanges mehr und mehr in den Vordergrund tritt.

### **Zusammenfassung**

Willensschwäche erscheint auf zwei Weisen: als eigene Willensschwäche und als die von anderen Personen. Welche Phänomene genau unter den Begriff der Willensschwäche fallen, bzw. was als Willensschwäche gelten soll, ist umstritten. Diese Untersuchung fügt sich in den Rahmen eines interdisziplinären Rahmenmodells ein, das zwischen drei Stufen des Handlungsprozesses unterscheidet: der Generierung von Handlungsoptionen, der Evaluation und Selektion der Optionen und deren Ausführung. In dieser Untersuchung konzentriere ich mich auf Willensschwäche im klassischen Sinne, also als Handeln wider das bessere Urteil, das der Phase zwei zuzuordnen ist, der Evaluation und Selektion von Handlungsoptionen, werde aber auch auf Willensschwäche als vorschnelle Änderung bereits bestehender Intentionen eingehen, das Phase drei zuzuordnen ist. Statt den Bezeichnungen ‚Akrasia‘ und ‚Willensschwäche‘ verwende ich

‚Willensschwäche‘ als Überbegriff für zwei Arten von Phänomenen: *motivationale* Willensschwäche (Handeln wider das bessere Urteil) und *exekutive* Willensschwäche (Handeln wider eigene Pläne). Statt der Ausdehnung des Begriffs der Willensschwäche auf theoretische Fehlleistungen argumentiere ich gegen die Existenz willensschwacher Überzeugungen und damit für eine gewisse Asymmetrie zwischen theoretischer und praktischer Rationalität. Willensschwäche ist zudem ein komplexer Begriff der auf andere Begriffe verweist: Evaluation, Intention, Verantwortung und Normativität. Diese Komponenten bilden die philosophischen Problemfelder, die für sich und in ihrem Zusammenhang hier untersucht werden. Zumindest graduell lässt sich Willensschwäche von anderen psychischen und psychopathologischen Phänomenen abgrenzen. Dies betrifft vor allem den Glauben der Person in die eigene Handlungsfreiheit sowie das generelle affektive Niveau der Person.

## II Willensschwäche als philosophisches Problem

In diesem Kapitel möchte ich in die philosophische Diskussion über Willensschwäche anhand zweier paradigmatischer Denker einführen. Aristoteles und Donald Davidson fungieren dabei nicht nur vorbildlich im Sinne der Klarheit und Prägnanz ihrer Untersuchungen zu Willensschwäche, ihre verschiedenen Ansätze verdeutlichen zudem die Verschiedenheit einer moralphilosophischen und handlungstheoretischen Perspektive auf das zugrundeliegende Phänomen. Die Perspektive entscheidet dabei über den Zugriff auf das Phänomen, die Rahmung, sowie die Einordnung in Problemfelder und die Signifikanz, welche die verschiedenen Probleme, die mit Willensschwäche verbunden sind, gewinnen. Für Aristoteles steht Willensschwäche nicht in erster Linie als Phänomen einzelner willensschwacher Handlungen im Vordergrund, sondern als moralisch bedeutsamer Charakterzug, der sich in willensschwachen Handlungen äußert. Die willensschwache bzw. akratische Person ist eine, die über ein Verständnis des guten Lebens verfügt, aber in ihren konkreten Handlungen, ihrem Verhalten und Streben, ihrem eigenen Verständnis nicht gerecht wird. Die Kritikwürdigkeit ergibt sich aus diesem Verfehlen des Idealen, das nicht subjektiv bestimmt ist als bloße Meinung über das Erstrebenswerte, sondern objektiv als geteilte Auffassung des ideal Erstrebenswerten. Hinter diesem Anspruch, Willensschwäche nicht bloß auf einzelne Akte, sondern als problematischen Lebensentwurf zu charakterisieren, treten andere Aspekte der Detailanalyse willensschwacher Handlungen zurück. Bei Davidson wird von vornherein auf die moralphilosophische Rahmung des Phänomens der Willensschwäche verzichtet. Das Hauptaugenmerk liegt auf willensschwachen Handlungen, deren Intentionalität, Erklärung und Begründung. Die Kritikwürdigkeit der Willensschwäche basiert auf einem minimal rationalen Anspruch der Kohärenz, der Übereinstimmung bestimmter evaluativer Urteile und deren konkreter Ausführung. Der Inhalt dieser Urteile spielt für die Analyse keine Rolle, alleine die logische Struktur in Gemeinschaft mit grundlegenden Prinzipien rationalen Handelns soll die normative Signifikanz von Willensschwäche tragen. Trotz der Verschiedenheit der beiden Zugriffe auf unterschiedlichen Ebenen – metaphysisch, epistemologisch, methodisch – treten dieselben Problemfelder in beiden Ansätzen auf, wenn auch in verschiedener Gewichtung. Die Darstellungen sollen vor allem das Ziel, eine Problem- und Diskussionsexposition zu geben, erfüllen, die interpretatorische Genauigkeit muss hinter dieses Ziel angesichts der umfangreichen und im Falle Aristoteles sicher nicht mehr zu überblickenden Diskussionsgeschichte zurücktreten.<sup>27</sup>

---

<sup>27</sup> Für dieses Vorgehen vgl. die Unterscheidung von ‚traditional classical scholarship‘ und ‚philosophical scholarship‘ bei Charles 1984.

## II.1 Akrasia bei Aristoteles

Im Folgenden möchte ich die aristotelische Position in Bezug auf Akrasia<sup>28</sup> untersuchen. Dies ist deshalb relevant, da hier die wesentlichen philosophischen Probleme mit Willensschwäche erstmals eingehend und differenziert diskutiert worden sind, Aristoteles die weitere Diskussion entscheidend geprägt hat und seine Position noch heute vielen attraktiv erscheint. Ich gehe in folgenden Schritten vor: Zuerst werde ich die aristotelische Diskussion der Akrasia in den Kontext seiner ethischen Schriften, v.a. der Nikomachischen Ethik (NE), einbetten (1). Danach werde ich Aufbau und Struktur des VII. Buches der NE darstellen (2). Darauf aufbauend werde ich einige Eckpunkte der aristotelischen Analyse der Akrasia herausarbeiten (3), die wenig umstritten sind, und mich dann den umstrittenen Teilen zuwenden. Umstritten ist vor allem die genaue Interpretation des Wissens, über das der Akratiker nach Aristoteles verfügt (4).

### II.1.1 Die Diskussion der Akrasia innerhalb der NE

Aristoteles widmet sich der philosophischen Analyse der Akrasia ausführlich in der ersten Hälfte des siebten Buches<sup>29</sup> seiner *Nikomachischen Ethik*.<sup>30</sup> In den Büchern zuvor hat Aristoteles seine Vorstellung einer Ethik entwickelt, in deren Zentrum die Frage nach dem guten Leben steht. In Buch I schließt er aus der Zielhaftigkeit alles menschlichen Handelns, seines teleologischen Charakters, auf die Notwendigkeit, letzte Ziele anzunehmen, die nicht wieder nur als Mittel zu einem übergeordneten Zweck verfolgt werden (NE I.1). Dieses letzte Ziel bestimmt er formal als das beste Gut (*ariston*), und identifiziert es für den Menschen als *eudaimonia* (Glück): die

---

<sup>28</sup> Der griechische Begriff ‚Akrasia‘ ist gebildet aus dem Alpha privativum (als Ausdruck der Negation) und ‚kratos‘ (Macht, Kraft) und bedeutet demnach angewendet auf Akteure ursprünglich so viel wie ‚fehlende Selbstkontrolle‘ oder ‚Unbeherrschtheit‘. Übersetzungsvarianten: Unenthaltbarkeit; lat. *incontinentia*; engl. *lack of self-control*, *incontinence*. Gegenteil: *enkrateia*; Beherrschtheit; Selbstbeherrschung; Enthaltbarkeit; lat. *continentia*; engl. *continence*; *self-control*. Vgl. Wolf 2007, 257; 260; Bostock 2000, 123.

<sup>29</sup> Der zweite Teil des siebten Buches (NE VII.12–15) widmet sich der Analyse von Lust und Unlust. Die Einteilung der Nikomachischen Ethik in Bücher und Kapitel stammt nicht von Aristoteles selbst, sondern von seinen Editoren. Ich greife auf sie zurück, da sie nützlich ist und dabei hilft, den komplexen Text in übersichtliche sinnvolle Einheiten zu gliedern. Die übliche Einteilung im englischsprachigen Raum weicht dabei von der im deutschsprachigen etwas ab. So entspricht VII.5 bei der Ausgabe von Ursula Wolf (NE) dem Kapitel VII.3 bei der Ausgabe von Christopher Rowe und Sarah Broadie (Aristoteles 2002). Zur Einteilung der Kapitel vgl. Natali 2009a, 1. Wir wissen nicht, wer ursprünglich die disparaten Behandlungen der Akrasia und der Lust in ein Buch zusammengestellt hat. Eine Erklärung ist, dass sie aufgrund ihrer Kürze beide auf einen Papyrus passen. Unklar ist auch die Identität des Editors und von ‚Nikomachus‘ – ist dies eine Widmung an den Sohn des Aristoteles mit gleichem Namen, oder doch der Name des Editors der Schrift?

<sup>30</sup> Das Buch VII der Nikomachischen Ethik ist identisch mit Buch VI der Eudemischen Ethik. Zum Verhältnis der beiden Ethiken vgl. Kenny 1978; Wolf 2007. Vor allem ihre zeitliche Reihenfolge ist noch immer umstritten, wobei natürlich die gemeinsamen Bücher im Mittelpunkt dieser Diskussion stehen. So ist strittig, ob die gemeinsamen Bücher ursprünglich zu NE oder zu EE gehören. Die dritte ethische Abhandlung, die *Magna Moralia*, stammt nach der üblichen Forschungsmeinung nicht von Aristoteles. Die vierte Abhandlung *Über Tugenden und Laster* stammt nach einhelliger Meinung nicht von Aristoteles. Ich werde beide aus diesen Gründen hier nicht berücksichtigen. Vgl. auch Aristoteles 2005, xi.

eudaimonia sei das letzte Ziel, auf das sich schließlich alles Wollen und Streben der Menschen richte (NE I.2). Ziel und Zweck der restlichen Analyse ist es nun, die eudaimonia für den Menschen auch inhaltlich näher zu bestimmen, d.h. zu zeigen, worin die spezifische eudaimonia für den Menschen besteht. Als Ansatz und Ausgangspunkt verwendet Aristoteles hierfür das umstrittene Ergon-Argument (NE I.6):<sup>31</sup> Für jeden Gegenstand, der eine bestimmte Funktion (ergon) besitzt, können wir angeben, was es heißt, diese Funktion gut zu erfüllen. So ist ein Messer gut, wenn es gut schneidet, ein Arzt gut, wenn er Menschen gut heilen kann usw. Wenn nun der Mensch auch allgemein eine solche Funktion hätte, dann ließe sich über die Analyse dieser Funktion angeben, worin das Glück als bestes Gut für den Menschen bestehe, da diese Funktion gut zu erfüllen zugleich hieße, ein gutes Leben zu führen, und ein gutes Leben zu führen hieße, eudaimon bzw. glücklich zu sein. Die Gutheit bzw. Tugend (arete) eines menschlichen Lebens liegt also darin, die spezifisch menschliche Funktion, sein ergon, möglichst gut und abschließend zu verwirklichen. Menschen unterscheiden sich nach Aristoteles von anderen Lebewesen durch ihre Vernunftfähigkeit: die Betätigung dieser Fähigkeit ist die spezifisch menschlich Weise zu leben, und möglichst gut lebt deshalb, wer sein Leben in Betätigung dieser Fähigkeit vollzieht, d.h. vernünftig lebt. Vernünftig kann ein Mensch jedoch nach Aristoteles auf zumindest zwei Arten leben, nämlich einerseits hinsichtlich einer vernünftigen Formung seiner affektiven Strebungen (Tugenden des Charakters oder ethische Tugenden), andererseits in der Betätigung der reinen Vernunfttätigkeit selbst (Tugenden des Denkens oder dianoetische Tugenden, NE I.13). Diese Zweiteilung ergibt sich aus einer Dreiteilung der menschlichen Psyche in einen vernunftlosen (threptikon oder vegetativ), einen affektiv-strebenden, jedoch der Vernunft zugänglichen (orektikon oder Strebevermögen) und einen vernünftigen Teil (logos).<sup>32</sup> Ein Großteil der Nikomachischen Ethik widmet sich der Analyse der spezifisch menschlichen arete oder Gutheit im Sinne der Betätigung der Vernunft im Bereich des Strebevermögens bzw. des Charakters.

In Buch II bestimmt Aristoteles die Gutheit (arete) des Charakters allgemein und abstrakt als Disposition (hexis), als feste Verfassung des Charakters oder Haltung gegenüber dem Strebevermögen, die sich in Vorsätzen äußert und in einer Mitte zwischen zwei extremen

---

<sup>31</sup> ‚Ergon‘ kann übersetzt werden mit: Funktion, eigentümliche Leistung, eigentümliche menschliche Tätigkeit; engl. function. Vgl. Wolf 2007, 260. Zum Argument vgl. Müller 2003; Ricken 1976, 21–34; Wolf 2007, 23–56.

<sup>32</sup> Corcilius 2008a vertritt die Ansicht, dass das Strebevermögen kein eigener Seelenteil sei, sondern lediglich der Art nach etwas, das bewegt. „Zusammengenommen heißt dies, dass Aristoteles die Strebefähigkeit nicht als einen genuinen Teil der *Seele* angesehen haben kann.“ (Corcilius 2008a, 50) D.h. das Strebevermögen ist kein genuines Vermögen der *Seele selbst*, sondern Vermögen von belebten (beseelten) Körpern.

Haltungen liegt, wie sie der Kluge (*phronimos*) in einer Überlegung bestimmen würde.<sup>33</sup> Der genaue Sinn dieser Lehre von der Tugend als rechter Mitte zwischen zwei entgegengesetzten Lastern ist umstritten.<sup>34</sup> Klar ist, dass sich Aristoteles bemüht, die rechte Mitte in spezifischen Affekt- und Handlungsbereichen zu analysieren, also konkrete Tugenden und ihre entsprechenden Laster zu benennen und zu analysieren. Dies geschieht in den Büchern II bis V. Das Buch IV ist ganz der Analyse einzelner Tugenden gewidmet, während sich das Buch V ausschließlich auf die Tugend der Gerechtigkeit konzentriert. Buch III widmet sich neben einzelnen Tugenden v.a. Fragen der Vorsätzlichkeit und Kriterien der Zurechenbarkeit von Handlungen (NE III.1–8). Buch VI ist der Analyse der Vernunfttugenden (*dianoetische* Tugenden) gewidmet. Hier geht es vor allem um das Verhältnis der Tugend der Klugheit (*phronesis*) zu den übrigen Charaktertugenden sowie dessen Abgrenzung zu anderen Vernunfttugenden. Buch VII ist wie erwähnt der Analyse der Unbeherrschtheit gewidmet (NE VII.1–11) sowie im zweiten Teil einer ersten Analyse der Lust (NE VII.12–15). Die Bücher VIII und IX handeln von dem Phänomen der Freundschaft und dessen Bedeutung und Beitrag für ein gutes Leben. Buch X nimmt zuerst die Analyse der Lust aus einem anderen Blickwinkel wieder auf (NE X.1–5), um danach die *eudaimonia* im politischen Leben mit derjenigen im theoretischen Leben zu vergleichen (NE X.6–10).

### II.1.2 Aufbau von NE VII.1–11

Die Untersuchung lässt sich grob in drei Teile gliedern: 1) Vorläufige Bestimmung des Gegenstandes und der angemessenen Methode der Untersuchung (NE VII.1) sowie eine Auflistung der anerkannten Meinungen über den Gegenstand (NE VII.2); 2) Auflistung der Schwierigkeiten (*aporiai*), die mit dem Gegenstand und den anerkannten Meinungen verbunden sind (NE VII.3–4); 3) Erörterung und Lösung einiger Schwierigkeiten (NE VII.5–11).

1) Als Gegenstand der Untersuchung (NE VII.1) bestimmt Aristoteles *Akrasia* zunächst als eine grundlegende Verfassung des Charakters (*hexis*), die es neben Schlechtigkeit (*kakia*) und tierischer Rohheit (*theriotes*) zu meiden gilt.<sup>35</sup> Deren Gegensätze sind *Enkrateia* oder Beherrschtheit, charakterliche Gutheit (*arete*) und göttliche (übermenschliche) Tugend. Neben *Akrasia* nennt Aristoteles auch Weichlichkeit (*malakia*) und dessen Gegenteil, die Ausdauer (*karteria*), als Gegenstände, die untersucht werden müssen, da sie anscheinend zusammengehören. Wie sich herausstellen wird, ist Beherrschtheit und Unbeherrschtheit primär auf Lust, Weichlichkeit und

---

<sup>33</sup> „Die Tugend ist also eine Disposition (*hexis*), die sich in Vorsätzen äußert (*prohairesis*), wobei sie in einer Mitte liegt, und zwar der Mitte in Bezug auf uns, die bestimmt wird durch die Überlegung (*logos*), das heißt so, wie der Kluge (*phronimos*) sie bestimmen würde.“ (NE 1106 b36)

<sup>34</sup> Vgl. Wolf 2007, 71ff.; Wolf 2010; Urmson 1980.

<sup>35</sup> Vgl. Cooper 2009, 12. Zur tierischen Rohheit (engl. beastliness) vgl. Cooper 2009, 17f; Natali 2009b.

Ausdauer dagegen primär auf Unlust bezogen. Im Zentrum der Untersuchung steht jedoch Unbeherrschtheit.<sup>36</sup>

Sein methodisches Vorgehen beschreibt Aristoteles so, dass er zuerst alltägliche Überzeugungen und anerkannte Meinungen (*ta phainomena*,<sup>37</sup> *legomena* und *endoxa*) über *Akrasia* sammelt, sodann deren Schwierigkeiten untereinander bestimmt und diese Schwierigkeiten schließlich so gut es geht aufzulösen versucht (NE VII.2).<sup>38</sup>

„Wir müssen nun, wie auch sonst, zuerst darlegen, was über die Gegenstände wahr zu sein scheint (*ta phainomena*), und die Schwierigkeiten durchgehen (*diaporein*), um dann auf diese Weise, wenn möglich, die Wahrheit aller anerkannten Meinungen (*endoxa*) über diese Affektionen (*pathe*) nachzuweisen, oder wenn nicht, [wenigstens] die Wahrheit der meisten und wichtigsten Meinungen. Denn wenn wir die Schwierigkeiten auflösen und die anerkannten Meinungen bestehen bleiben, dann wird der Gegenstand ausreichend geklärt sein.“ (NE 1145 b3ff. Hinz. v. Ursula Wolf)

Aristoteles zählt sechs anerkannte Meinungen (NE VII.2) über *Akrasia* auf, die im Zuge der Untersuchung zu berücksichtigen sind. Es sind dies:<sup>39</sup>

M1) Beherrschtheit und Ausdauer sind gut und lobenswert, Unbeherrschtheit und Weichlichkeit dagegen schlecht und tadelnswert.

M2) Der Beherrschte bleibt bei seiner Überlegung, der Unbeherrschte nicht.

M3) Der Unbeherrschte weiß, dass es schlecht ist, was er tut, und tut es dennoch aufgrund des Affektes. Der Beherrschte hingegen handelt gemäß seinem Wissen.

M4) Einige identifizieren Mäßigkeit<sup>40</sup> (*sophrosyne*) mit Beherrschtheit und Unmäßigkeit mit Unbeherrschtheit, andere nicht.

M5) Einige meinen, kluge Menschen (*phronimoi*) können nicht unbeherrscht sein, andere hingegen meinen, dass dem nicht so ist.

M6) Unbeherrscht werden Menschen nicht nur in Bezug auf sinnliche Begierden, sondern auch in Bezug auf Zorn, Ehre und Gewinn genannt.

---

<sup>36</sup> „Thus, lack of control is declared as the centre of his interest in this little treatise, and self-control is mentioned only secondarily as part of the subject making up his new topic, because it is uncontrol’s opposite.” (Cooper 2009, 13)

<sup>37</sup> ‚*Ta phainomena*‘ ist zweideutig und kann einerseits anerkannte Meinungen über Erscheinungen, andererseits die Erscheinungen selbst meinen. Vgl. NE 373 Fn. 7. Der Kontext legt jedoch nahe, dass Aristoteles hier auf Meinungen über Erscheinungen verweist, die er anschließend selbst anführt. Vgl. Cooper 2009, 22f.

<sup>38</sup> „Aristotle’s proposed method, then, consists in collecting and setting down a series of things that are said (or just assumed by being thought), whether in ordinary life or by philosophers or other notable writers and thinkers, specifically on the topic of lack of control and self-control – that is, about what these conditions *are*.” (Cooper 2009, 27) Über das methodische Vorgehen vgl. Wolf 2007, 57ff.; Bostock 2000, 214–235. Allgemeiner vgl. auch Höffe 2010a; Müller 2009, 124–130. Zur endoxastischen Methode vgl. auch Barnes 1980.

<sup>39</sup> Vgl. Wolf 2007, 166; Cooper 2009, 21ff.

<sup>40</sup> Vgl. NE 1117 b22ff.

Die anerkannten Meinungen umfassen sowohl die alltägliche moralische Beurteilung von Akrasia (M1), Fragen der korrekten begrifflichen Explikation (M2, M3, M5) als auch Fragen der Abgrenzung von anderen Phänomenen (M4) sowie die Unterscheidung von eigentlichen und uneigentlichen Zuschreibungen von Akrasia (M6).

2) Nach Auflistung der anerkannten Meinungen wirft Aristoteles sechs Schwierigkeiten (NE VII.3–4) mit dem Phänomen der Akrasia auf:<sup>41</sup>

S1) Wie kann jemand ein richtiges Urteil haben und dennoch entgegen diesem Urteil handeln? (Problem mit M2 und M3 sowie handlungstheoretischen Prämissen)

S2) Wie können sowohl Beherrschtheit als auch Mäßigkeit gelobt werden, wenn sich beide Haltungen auf Lust beziehen, der Beherrschte aber übermäßige Begierden hat, der Mäßige nicht? (Problem mit M1 und M4)

S3) Wenn Beherrschtheit darin bestünde, bei jedem *beliebigen* Urteil zu bleiben, dann wäre die Beherrschtheit schlecht und die Unbeherrschtheit gut, wenn es sich dabei um ein falsches Urteil handelt.<sup>42</sup> (Problem mit M1 und M2)

S4) Unverstand und Unbeherrschtheit könnten gemeinsam eine Tugend ergeben: Jemand könnte gegen sein Urteil handeln (aus Unbeherrschtheit), das jedoch falsch ist (aus Unverstand), und damit gerade das richtige tun. (Problem mit M1, M2 und M5)

S5) Unmäßigkeit scheint schlechter zu sein als Unbeherrschtheit, da der Unmäßige vorsätzlich den Begierden folgt, der Unbeherrschte nicht. Doch scheint der Unmäßige auch besser zu sein als der Unbeherrschte, da er leichter zu heilen ist, indem man ihn von der Falschheit seiner Auffassung überzeugen kann, während der Unbeherrschte schon vom Richtigen überzeugt ist und dennoch nicht entsprechend handelt. (Problem mit M1, M3 und M4)

S6) Wenn Unbeherrschtheit nicht nur im Bereich der sinnlichen Begierden auftritt, sondern auch in anderen Bereichen, was ist der eigentliche Bereich der Unbeherrschtheit bzw. wen nennen wir unbeherrscht überhaupt (*haplos*), und nicht bloß hinsichtlich eines spezifischen Bereiches? (Problem mit M6)

Es fällt auf, dass Aristoteles die erste Schwierigkeit – die Wissens-Aporie – im Vergleich zu den anderen Schwierigkeiten sehr ausführlich einführt. Dies sollte jedoch nicht so interpretiert werden, als würde Aristoteles zwei unabhängige Lösungen für das Problem des Wissens des Akratikers geben – eine hier, die andere später (NE VII.5). Bei der Einführung des Problems

---

<sup>41</sup> Vgl. Wolf 2007, 167f.; Cooper 2009, 29ff.

<sup>42</sup> Dieses Problem wird in der heutigen Zeit unter dem Titel ‚inverse Akrasia‘ verhandelt, wobei hier jedoch im Unterschied zur NE gefragt wird, ob willensschwaches Handeln auch rational sein kann. Vgl. Kap. VI.1.2.

orientiert er sich vielmehr an der sokratischen Problembestimmung, die jedoch in Folge nicht der eigenen Problemformulierung entspricht.<sup>43</sup>

In NE VII.4 verkleinert Aristoteles diese Liste an Schwierigkeiten nochmals und gibt an, in welcher Reihenfolge er diese Probleme diskutieren wird:

„Ungefähr von dieser Art sind also die Schwierigkeiten, die sich ergeben. Einige von ihnen müssen aufgelöst werden, die anderen muss man auf sich beruhen lassen. Denn das Auflösen der Schwierigkeit ist ein Finden [der Wahrheit]. Zuerst müssen wir untersuchen, ob die Unbeherrschten wissend (*eidos*) handeln oder nicht und in welcher Weise wissend. Zweitens ist zu fragen, welches die Gegenstände sind, in Bezug auf die der Unbeherrschte und der Beherrschte anzusiedeln sind, das heißt, ob jede Lust (*hedone*) und Unlust (*lype*) oder [nur] einige bestimmte Arten. Weiter, ob der Beherrschte (*enkratēs*) und der Ausdauernde (*kearterikos*) derselbe sind oder verschieden. Ähnlich müssen wir auch die anderen Fragen behandeln, die mit dieser Betrachtung zusammenhängen.“ (NE 1146 b6–12)

Bereits hier ist festzuhalten, dass Aristoteles zwar nicht alle der vorgestellten Schwierigkeiten explizit zu lösen versucht, aber auch nicht alle Schwierigkeiten, die er tatsächlich diskutiert, in seine Liste an Schwierigkeiten aufgenommen hat.<sup>44</sup> Dennoch lassen sich die Lösungen zu den Schwierigkeiten aus seinen Diskussionen erschließen, auch wenn er nicht immer direkt auf diese zurückverweist.

3) In Folge (NE VII.5–11) erörtert Aristoteles einige der genannten Schwierigkeiten und versucht diese zu lösen. Zuerst beschäftigt er sich ausführlich mit der Frage, ob und in welchem Sinne der Unbeherrschte wissend handelt (NE VII.5). Danach klärt er in einem ersten Anlauf die Frage, auf welchen Bereich die Unbeherrschtheit bezogen ist und wer in einem absoluten und wer nur in einem übertragenen Sinne als unbeherrscht zu bezeichnen ist (NE VII.6). Dann diskutiert er die Frage, ob Unbeherrschtheit im eigentlichen Sinne schlechter ist als Unbeherrschtheit im übertragenen Sinne bezogen auf den Affekt des Zorns (*thymos*) (NE VII.7). In NE VII.8 unterscheidet er verschiedene Formen von Unbeherrschtheit und diskutiert die Unterscheidung von Unbeherrschtheit und Weichlichkeit. In NE VII.9 untersucht er die Frage, ob Unmäßigkeit schlechter ist als Unbeherrschtheit. NE VII.10 diskutiert die Frage, ob sich Beherrschtheit auf jedes beliebige, oder nur auf richtige Urteile bezieht. NE VII.11 diskutiert schließlich die Frage, ob der Mäßige beherrscht ist und ob der Kluge auch unbeherrscht handeln kann.

### II.1.3 Interpretatorische Eckpunkte

Angesichts einiger divergierender Interpretationen bietet es sich an, weniger kontroverse Eckpunkte der aristotelischen Auffassung von Akrasia von den strittigen zu trennen. Bezüglich

---

<sup>43</sup> Zu diesen Punkt vgl. Kap. II.1.4 sowie Corcilius 2008b, 144–147. Aber: Cooper 2009, 36f.

<sup>44</sup> Kritisch beurteilt dies Cooper: „We can see, then, that despite the superficial clarity and decisiveness of Aristotle’s account in chapter 1 of the method he intends to follow, we do not find him actually setting out first (in chapter 2) all the puzzles (or anyhow all the questions and problems for investigation) that he intends to try to resolve. Nor does he in fact solve (in chapters 3–10) all the ones he does set out there.“ (Cooper 2009, 38)

der anerkannten Meinungen, die Aristoteles zu Beginn seiner Untersuchung auflistet, lässt sich folgendes feststellen: M1 bleibt vollauf erhalten. M2 und M3 je nach Interpretation mehr oder weniger. Bei M4 stellt sich Aristoteles auf die Seite derer, die Mäßigkeit, Beherrschtheit, Unmäßigkeit und Unbeherrschtheit als voneinander getrennte Phänomene behandeln, gibt aber einer Erklärung, weshalb manche diese Unterschiede verneinen.<sup>45</sup> Bei M5 verteidigt Aristoteles die Meinung, dass ein kluger Mensch (*phronimos*) nicht unbeherrscht handeln könne, gibt aber auch hier eine Erläuterung, weshalb manche dies übersehen.<sup>46</sup> Auch M6 stimmt Aristoteles zu, indem er eine eigentliche Zuschreibung von *Akrasia* von uneigentlichen unterscheidet und erstere auf den Bereich sinnlicher Begierden beschränkt.

Nun zu den Eckpunkten:

E1) Unbeherrschtheit ist eine Verfasstheit des Charakters, die zu tadeln ist.<sup>47</sup> Dennoch ist Unbeherrschtheit kein Laster und Beherrschtheit keine Tugend.<sup>48</sup> Dafür spricht erstens, dass Aristoteles sie nicht in seine Liste der Tugenden und Laster in NE II.7 aufgenommen hat. Dafür spricht zweitens auch, dass die Mühe, Unbeherrschtheit vom Laster der Unmäßigkeit zu unterscheiden, vergeblich wäre, wenn beide zwar auf denselben Bereich bezogen (sinnliche Begierden), aber dennoch nicht voneinander verschieden wären. Dagegen spricht, dass Beherrschtheit gelobt wird, und man deshalb meinen könnte, sie sei daher auch eine Tugend. Aber nicht alles, was zu Recht gelobt wird, ist eine Tugend. Nur umgekehrt ist jede Tugend auch lobenswert. (Bestätigung von M1)

E2) Aristoteles argumentiert offensichtlich für eine zumindest vierteilige Unterscheidung zwischen Mäßigkeit, Unmäßigkeit, Beherrschtheit und Unbeherrschtheit. Gemeinsam ist allen diesen Phänomenen, dass sie sich im eigentlichen (ursprünglichen) Sinne auf denselben Bereich beziehen, nämlich den Bereich sinnlicher Lust. Kompliziert wird diese Vierteilung durch den Umstand, dass die Mäßigkeit als Einzeltugend die richtige Mitte zwischen zwei Lastern bildet. Es gibt demnach zwei Formen der Unmäßigkeit, nämlich *akolasia* und *anaesthesia* (Stumpfheit). Es gibt zwar auch zwei Arten von Unbeherrschtheit (überlegte und unüberlegte bzw. Schwäche und

---

<sup>45</sup> Diese lassen sich von den Ähnlichkeiten der Phänomene untereinander dazu verführen, sie nicht voneinander abzugrenzen.

<sup>46</sup> Seine Unterscheidung zwischen Geschicklichkeit (*deinotēs*) und Klugheit (*phronēsis*). Vgl. NE 1144 a23–b17

<sup>47</sup> Vgl. Cooper 2009, 10.

<sup>48</sup> „Unbeherrschtheit und Beherrschtheit fallen nicht unter die Begriffe von Tugend und Laster, sondern konstituieren moralische Zustände *sui generis*.“ (Müller 2009, 114) „Auch die Beherrschtheit (*enkrateia*) ist keine Tugend, sondern eine gemischte Disposition.“ (NE 1228b 33–34) „Tatsächlich aber ist Tugend und Beherrschtheit verschieden.“ (EE 1227b 17) Aber: „Die Unbeherrschtheit wird nicht nur als Fehler (*hamartia*), sondern auch als eine Art von Laster (*kakia*) getadelt, entweder überhaupt oder in einem Teil, während das für keine der genannten besonderen Arten von Unbeherrschtheit der Fall ist.“ (NE 1148 a2–4) Vgl. auch EE 1223b 11–12. Teun Tieleman siedelt Beherrschtheit als rechte Mitte zwischen Unbeherrschtheit auf der einen Seite und einem namenlosen Laster (‘under-indulgence’) auf der anderen Seite an. Vgl. Tieleman 2009, 174ff.

Voreiligkeit), doch sind diese nicht analog zu zwei Laster bezogen auf eine Tugend (Beherrschtheit) zu verstehen. Beherrschtheit ist eine lobenswerte Eigenschaft von Personen, deren Affekte und Begierden noch nicht völlig mit ihren Einsichten über das tatsächlich Gute übereinstimmen und dennoch aufgrund dieser Einsichten handeln. Sie sind keine *phronimoi*, aber auch keine lasterhaften Menschen. (Klärung M4 und Lösung von S2)

E3) Der Kluge (*phronimos*) kann nicht unbeherrscht handeln, und der Mäßige ist nur in einem übertragenen Sinne beherrscht. Aristoteles weist explizit darauf hin, dass niemand zugleich klug und unbeherrscht sein könne.<sup>49</sup> Der Mäßige ist deshalb nicht beherrscht, weil er keine entsprechend übermäßigen Begierden hat, die er beherrschen müsste. Da der Beherrschte wie der Mäßige jedoch das richtige tun, könnte man dazu verleitet werden, diese beiden Charakterdispositionen nicht voneinander zu unterscheiden. (Klärung von M5)

E4) Nach Aristoteles besteht eigentliche Unbeherrschtheit darin, nicht gegen ein beliebiges, sondern gegen ein richtiges bzw. wahres Urteil zu handeln.<sup>50</sup> Dies bedeutet nicht, dass das Handeln gegen ein vermeintlich richtiges Urteil nicht dieselben handlungstheoretischen Probleme verursacht wie Handeln wider das bessere Wissen. So grenzt sich Aristoteles in der Diskussion um die Zuschreibung von Wissen bewusst von Positionen ab, die vorgeben, das handlungstheoretische Problem der *Akrasia* zu lösen indem sie zwischen Wissen und bloßer Meinung unterscheiden.<sup>51</sup> Dennoch ist Willensschwäche für Aristoteles primär kein handlungstheoretisches Phänomen, sondern ein moralphilosophisches. Als solches muss klar sein, weshalb und wie es moralisch relevant ist. (Lösung von S3 und S4)

E5) Unmäßigkeit ist normativ schlechter zu bewerten als Unbeherrschtheit.<sup>52</sup> Der Unbeherrschte handelt gegen seinen Vorsatz, empfindet Reue über seine Handlungen und ist außerdem leichter heilbar als der Unmäßige. (Lösung von S5)

E6) Aristoteles unterscheidet Unbeherrschtheit im eigentlichen Sinne von Unbeherrschtheit im uneigentlichen Sinne anhand der Gegenstände bzw. affektiven Bereiche, auf die sie bezogen sind. Im eigentlichen Sinne unbeherrscht ist jemand nur bezogen auf körperliche Lust und Unlust.<sup>53</sup> Die zugrundeliegende Strebung ist demnach eine Begierde nach Lust bzw. die Vermeidung von Unlust. Im uneigentlichen oder übertragenen Sinne unbeherrscht ist jemand in Bezug auf Zorn

---

<sup>49</sup> „Es ist auch nicht möglich, dass ein und dieselbe Person zugleich klug (*phronimos*) und unbeherrscht ist. Denn wie gezeigt ist man gleichzeitig klug und gut im Charakter. Ferner: Klug ist man nicht nur durch Wissen, sondern auch durch die Disposition, dem Wissen entsprechend zu handeln.“ (NE 1152a 7–9) Wichtig ist in diesem Zitat auch die nachdrückliche Unterscheidung zwischen Klugheit und Wissen: Wissen alleine umfasst offensichtlich noch nicht die Disposition, dem Wissen gemäß zu handeln.

<sup>50</sup> Vgl. NE 1152 a7–13; NE 1151a 29–b4.

<sup>51</sup> Vgl. NE 1145b 21ff.; NE 1146b 24–30.

<sup>52</sup> Vgl. NE 1150b 29ff.

<sup>53</sup> Vgl. NE 1147b 31ff.

(bzw. Mut, Eifer oder andere Strebungen des thymos). Diese sprachliche Qualifizierung ist mit einem moralischen Bewertungsunterschied verbunden, da eigentliche Akrasia normativ schlechter bewertet wird als uneigentliche.<sup>54</sup> (Klärung von M6 und Lösung von S6)

E7) Aristoteles unterscheidet zwischen zumindest zwei Grundformen von Willensschwäche:

„Die Unbeherrschtheit ist teils Voreiligkeit (*propesteia*), teils Schwäche (*astheneia*). Denn die einen haben zwar überlegt, bleiben dann aber wegen des Affekts nicht bei dem, was sie überlegt haben, die anderen werden, weil sie nicht überlegt haben, vom Affekt geleitet.“ (NE 1150 b19–22)

Während in Fällen der Voreiligkeit die handelnde Person wegen des Affektes erst gar keine Überlegung über ihr Handeln anstellt, überlegt die Person in Fällen schwacher Unbeherrschtheit, weicht aber danach im Handeln von ihrer Überlegung ab.<sup>55</sup> Da Aristoteles diese Unterscheidung erst einführt, nachdem er das Problem des Wissens des Akratikers behandelt hat, ist strittig, wie die Unterscheidung zwischen Voreiligkeit und Schwäche in Bezug zum zentralen Kapitel NE VII.5 verstanden werden sollte: Behandelt Aristoteles hier beide Formen und die hintereinander, oder doch beide zugleich, oder gar nur eine von beiden?<sup>56</sup>

Einigermaßen Klarheit und Übereinstimmung herrscht also vor allem gegenüber Gegenstand und Bereich der Akrasia: Eigentliche Beherrschtheit und Unbeherrschtheit sind besondere Dispositionen des Charakters bezogen auf den Bereich sinnlicher Lust. Sie sind zu unterscheiden von den zu diesem Bereich gehörenden Lastern in Form von *akolasia* (Unmäßigkeit) und *anaesthesia* (Stumpfheit) und der Tugend der Mäßigkeit (*sophrosyne*). Die normative Bewertung dieser Dispositionen von Aristoteles ist auch klar: Tugendhaftigkeit vor Beherrschtheit vor Unbeherrschtheit vor Lasterhaftigkeit. Unbestritten ist auch, dass nach Aristoteles eigentliche Akrasia nicht darin besteht, gegen jedes beliebige Urteil, sondern wider das bessere *Wissen* zu handeln. Unklar ist vor allem, ob und in welchem Sinne nach Aristoteles dem Unbeherrschten Wissen über das richtige Handeln zugesprochen werden kann.<sup>57</sup> Unklar ist außerdem, welche Rolle seine Unterscheidung zweier Grundformen von Akrasia – Voreiligkeit und Schwäche – in seinem Konzept spielt.

#### **II.1.4 Das Wissen des Akratikers: Interpretationen von NE VII.5**

Ich möchte im Folgenden verschiedene Lesarten des zentralen Kapitels über die Frage, ob und wie der Akratiker wissend handelt, untersuchen. Grob gesehen lassen sich zwei interpretatorische

---

<sup>54</sup> Vgl. NE VII.6; Müller 2009, 111.

<sup>55</sup> Zur Unterscheidung zwischen Voreiligkeit und Schwäche vgl. auch NE 1152 a17–20.

<sup>56</sup> Vgl. Timmermann 2000.

<sup>57</sup> Diese Zusammenfassung soll keineswegs den Eindruck erwecken, als gäbe es außer dem Wissensproblem kein echtes Problem mit der aristotelischen Position. Probleme gibt es viele, dennoch erscheinen diese Eckpunkte als weithin geteilt und sehr gut durch direkte textliche Evidenzen abgesichert. Dies schließt weder einzelne abweichende Meinungen aus, noch Probleme des rechten Verständnisses dieser Eckpunkte.

Richtungen unterscheiden, eine Interpretation (Lesart A), die Aristoteles letztlich sehr nahe an der Position von Platons Sokrates im *Protagoras*<sup>58</sup> sieht, und eine welche dies nicht tut (Lesart B).<sup>59</sup> Im Kern drehen sich die verschiedenen Lesarten vor allem um die Frage, ob Aristoteles synchrone, klarsichtige Willensschwäche für möglich hält oder nicht, d.h. willensschwache Handlungen, bei denen der Akteur während der Handlung aktuell weiß, was am besten zu tun wäre, und es dennoch nicht tut.<sup>60</sup> Nach A ist diese Form der Willensschwäche unmöglich, nach B nicht. Dieser Unterschied liegt daran, dass nach A willensschwache Handlungen letztlich durch ein spezifisches kognitives Defizit erklärt werden. Wo dieses Defizit ausbleibt, da bleiben auch entsprechende willensschwache Handlungen aus. Mithin wäre es unmöglich, dass ein Akteur, der in einer Situation aktuell über das relevante Wissen, was in einer Situation zu tun sei, verfügt, willentlich anders handelt. Nach Lesart B gibt es neben Fällen von willensschwachem Handeln, bei denen die Akteure kognitive Defizite aufweisen, auch solche, in denen dies nicht zutrifft. Das kognitive Defizit erklärt nicht das willensschwache Handeln, sondern ist lediglich eine (mögliche) Begleiterscheinung der eigentlichen Ursache, d.i. ein übermäßiger Affekt. Da dieser Affekt auch die willensschwache Handlung des Akteurs verursachen kann, obwohl dieser aktuell weiß, was eigentlich in dieser Situation am besten zu tun wäre, ist klarsichtige synchrone Willensschwäche möglich. Eine entscheidende Rolle in dieser Diskussion betrifft die Interpretation des sogenannten ‚praktischen Syllogismus‘. Je nachdem, wie dessen Funktionsweise bestimmt wird, ändert sich auch die Interpretation der strittigen Stellen. Alleine diese Diskussion ist natürlich sehr intensiv geführt worden, weshalb ich mich hier nur auf das Wesentliche beschränken kann.<sup>61</sup> Vorausgeschickt sei, dass ich Positionen für plausibel halte, welche den praktischen Syllogismus lediglich als *Analogie* verstehen, um die Notwendigkeit, mit der bestimmte Bewegungen auf bestimmte Bedingungen folgen, zu veranschaulichen. Dies steht im Gegensatz zu Positionen, welche im praktischen Syllogismus vor allem eine *Illustration* praktischer Deliberation verstehen,

---

<sup>58</sup> Vgl. Platon 2009, 351b–358d; Kap. III.2.1.

<sup>59</sup> Zu dieser Einteilung vgl. Aristoteles 2002, 385–387; Bostock 2000, 123–142; Müller 2009, 124–141; Saarinen 1994, 8–19; Charles 2009; Price 2006; Dahl 1984. Ich vermeide hier bewusst die Bezeichnung dieser Lesarten als ‚traditionell‘ vs. ‚revisionistisch‘, da dies zwar historisch gesehen sachlich begründbar wäre, aber negativ konnotiert ist. Sachlich angemessen wäre auch die Unterscheidung in eine intellektualistische und eine gemäßigt-intellektualistische Lesart. Ich bleibe lieber bei der neutralen Unterscheidung zwischen Lesart A und B. Saarinen 1994 unterscheidet idealtypisch zwischen drei verschiedenen Interpretationen, welche als heuristische Modelle für die mittelalterliche Diskussion verwendet werden können: 1) Socratic-Aristotelian, 2) non-Socratic Aristotelian und 3) non-Aristotelian. „The first view is roughly characterized by the claim that there is something wrong in the premises of a practical syllogism which, in turn, prohibits the conclusion from being properly reached. According to the second view, the conclusion is grasped properly, and *akrasia* is due to the temporary neglect of this conclusion. The third view assumes a conception of autonomy in which the person’s will is something “more” than a rational desire alone.” (Saarinen 1994, 17f.) Diese Einteilung deckt sich nicht ganz mit der Einteilung in Lesart A und B.

<sup>60</sup> Hiermit hängt das Problem des Sokratismus bei Aristoteles zusammen, vgl. Müller 2009, 124ff.; Gosling 1990 38ff.

<sup>61</sup> Vgl. Corcilius 2008c für eine Übersicht über die Diskussion im 20. Jh.

die außerdem noch das Vorkommen menschlicher Handlungen erklären soll.<sup>62</sup> Ich beginne, indem ich Lesart A kurz darstelle. Im Anschluss daran nenne ich ein paar Probleme im Zusammenhang mit dieser Interpretation und mögliche Reaktionen. Es folgt eine Darstellung der Lesart B. Obwohl beide Lesarten auf textliche Evidenzen verweisen können und so jeweils plausible Alternativen der Interpretation darstellen, bin ich der Meinung, dass Lesart B angesichts der gesamten Konzeption der aristotelischen Ethik vorzuziehen ist.

### **a) Lesart A**

Bereits bei der Einführung des Problems, wie jemand wissentlich unbeherrscht sein kann, bezieht sich Aristoteles nach dieser Lesart<sup>63</sup> affirmativ auf die sokratische Position, wie sie im *Protagoras* überliefert ist:<sup>64</sup>

„Man könnte nun eine Schwierigkeit darin sehen, wie jemand ein richtiges Urteil haben [...] und dennoch unbeherrscht sein kann. Einige sagen, dass er sich unmöglich so verhalten könne, wenn er *Wissen* hat. Denn es wäre – wie Sokrates dachte – seltsam, wenn, obwohl Wissen (*episteme*) vorhanden ist, etwas anderes es beherrschen und wie einen Sklaven herumzerren würde. Sokrates lehnte ja die fragliche Auffassung vollständig ab in der Überzeugung, es gebe keine Unbeherrschtheit. Niemand, so pflegte er zu sagen, handelt gegen das Beste in der Überzeugung, dies zu tun, sondern [man handelt so nur] aufgrund von Unwissenheit (*di'agnoian*). Diese Aussage nun widerstreitet offensichtlich den Tatsachen (*phainomena*), und wir müssen Fragen über den Affekt des Handelnden stellen: Wenn er durch Unwissenheit entsteht, um welche Art der Unwissenheit handelt es sich? Denn dass der Unbeherrschte nicht meint, dass er so handeln solle, ehe er in den Affekt gerät, ist offensichtlich.“ (NE 1145b 21–31)

Aristoteles stimmt demnach Sokrates darin zu, dass *Akrasia* letztlich durch Unwissenheit entstehe.<sup>65</sup> Was er kritisiert ist, dass Sokrates nicht angegeben habe, um welche Art von Unwissenheit es sich genau handle und wie diese entstehe. Genau dieses Defizit versucht Aristoteles anschließend zu beheben. Als Beleg wird oft die zusammenfassende Feststellung von Aristoteles am Ende von NE VII.5 angeführt:<sup>66</sup>

„Da die zweite Prämisse eine Meinung über einen Wahrnehmungsgegenstand enthält und diese das Handeln bestimmt, besitzt der Unbeherrschte, da er sich im Affekt befindet, diese entweder gar nicht, oder er besitzt sie so, dass das Besitzen kein Wissen ist, sondern ein Sprechen wie das eines Betrunkenen, der Verse des Empedokles aufsagt. Und weil der letzte Begriff (*boros*) nicht allgemein ist und nicht in gleicher Weise Wissen zu enthalten scheint wie die allgemeine Prämisse, scheint sich auch das zu ergeben, was Sokrates suchte. Denn nicht, wenn das, was als Wissen im eigentlichen Sinn gilt, vorhanden ist, entsteht der Affekt, und es wird auch nicht dieses Wissen durch den Affekt hin- und hergezerrt, sondern wenn das wahrnehmende (*aisthetike*) Wissen vorliegt.“ (NE 1147b 9–17)

---

<sup>62</sup> Vgl. Corcilius 2008d. Ich halte seine Analyse des praktischen Syllogismus von Aristoteles für richtig.

<sup>63</sup> Für diese Lesart vgl. Robinson 1969; Price 2006; Bostock 2000, 123–142; Timmermann 2000; Hardie 1968.

<sup>64</sup> Aristoteles bezieht sich hier offensichtlich auf Platon 2009, 352b–c.

<sup>65</sup> Vgl. Bostock 2000, 124f.

<sup>66</sup> Vgl. auch „In der Tat verhält sich der Unbeherrschte nicht wie jemand, der weiß und betrachtet (*theorein*), sondern wie ein Schlafender oder Betrunkener. Und er handelt zwar wollend (*hekōn*) (denn in gewisser Weise weiß er, was er tut, und ebenso, zu welchem Zweck er es tut), ist aber nicht schlecht, da sein Vorsatz gut ist; er ist also halbschlecht.“ (NE 1152 a14–17) Diese Stelle wird demnach als Angabe eines kognitiven Defizits gedeutet.

Dies ist der grundlegende Rahmen für die Interpretation des einschlägigen Kapitels NE VII.5. Klaus Corcilius hat die sokratische Position anhand eines Dilemmas zweier Aussagen dargestellt:

- „(1) Wissen (επιστημῆ) von X beinhaltet Handeln in Übereinstimmung mit X
- (2) Akrasie ist ein Handeln wider besseres Wissen“ (Corcilius 2008b, 145)

Beide Aussagen können nicht zugleich auf einen Akteur zutreffen, da das Wissen des akkratischen Akteurs (2) zur entsprechenden Handlung laut (1) führen müsste. Sokrates löste dieses Dilemma auf, indem er das Phänomen der Akrasia anders deutete: „(2<sup>o</sup>) Akrasie ist ein Handeln aufgrund von Unkenntnis.“ (Corcilius 2008b, 145) Lesart A geht davon aus, dass auch Aristoteles im Grunde Satz (1) zustimmt, und nun versucht, durch eine Differenzierung der Wissensbedingung in Satz (2) zu zeigen, wie Akrasia möglich ist. Die aristotelische Aufgabe wäre also, die Art der Unkenntnis des Akkratikers so aufzuweisen, dass der Akkratiker sowohl wider besseres Wissen – die alltägliche Einschätzung des Phänomens der Willensschwäche – als auch aufgrund von Unkenntnis – die sokratisch-intellektualistische Handlungstheorie – handelt. Aristoteles rette also mit seiner Konzeption der Akrasia zugleich die Phänomene bzw. die als wahr erachteten Aussagen über diese (τα φαινόμενα) und die handlungstheoretische Grundprämisse von Sokrates.

Wie sieht nun diese Lösung laut Lesart A aus? Zuerst macht Aristoteles klar, dass das Problem nicht gelöst wird, wenn Akrasia als Handeln wider die bessere *Meinung* statt das eigene bessere *Wissen* bestimmt wird, also als: (2<sup>o</sup>) Akrasie ist ein Handeln wider die bessere Meinung.<sup>67</sup> Der subjektive Überzeugungsgrad kann in Fällen des bloßen Glaubens genauso hoch sein wie in Fällen tatsächlichen Wissens. Deshalb kann es keinen Unterschied hinsichtlich des handlungstheoretischen Problems machen, ob jemand wirklich weiß oder bloß glaubt, welche Handlung richtig wäre. Danach präsentiert Aristoteles seine eigene Lösung in vier aufeinanderfolgenden Schritten.

In Schritt 1<sup>68</sup> differenziert er zwei verschiedene Weisen, Wissen zu haben, eine dispositionelle und eine episodische. Ein Arzt weiß, wie er einen Menschen heilen kann, auch wenn er nicht gerade damit beschäftigt ist, jemanden zu heilen. Er besitzt das Wissen dispositionell und nicht aktuell. Analog könnte eine Person dispositionell wissen, was am besten zu tun sei, aber dieses Wissen in der relevanten Situation gerade nicht erwägen und daher auch nicht entsprechend handeln. Dennoch könnte man von dieser Person aussagen, dass sie wisse – nämlich dispositionell –, wie sie handeln sollte.

---

<sup>67</sup> Vgl. NE 1146 b24–31.

<sup>68</sup> Vgl. NE 1146 b31–36.

In Schritt 2<sup>69</sup> differenziert Aristoteles zwei verschiedene Arten von Wissensgehalten: allgemeine und partikuläre. Eine Person könnte etwas Allgemeines wissen, aber dieses Wissen nicht auf die einzelne Situation hin anwenden. Handelt er dann entsprechend, so würde er einerseits gegen sein eigenes besseres Wissen über das Allgemeine handeln, andererseits nicht, da er über das entsprechende partikuläre Wissen nicht verfügt oder es in der Situation nicht aktualisiert bzw. erwägt. Zudem können sowohl die allgemeinen als auch die partikulären Wissensgehalte aus mehreren Teilen bestehen, sodass auch hier noch weiter differenziert werden könnte, in Bezug worauf ein Akteur unwissend sein könnte. Lesart A sieht bereits hier Aristoteles Gebrauch von der Figur des praktischen Syllogismus machen und identifiziert Wissen über Allgemeines mit den Obersätzen, Wissen über weniger Allgemeines und Partikuläres mit Untersätzen solcher praktischen Syllogismen.

In Schritt 3<sup>70</sup> differenziert Aristoteles nun zwischen zwei verschiedenen Weisen, über dispositionelles Wissen zu verfügen. Ein Tänzer, der schläft, ist verschieden von einem Tänzer, der wach ist, aber gerade frühstückt und daher nicht tanzt. Beide besitzen dispositionelles Wissen, jedoch auf verschiedene Weise. Der eine muss, bevor er tanzen kann, zuerst aufwachen, während der andere jederzeit anfangen könnte zu tanzen.<sup>71</sup> Aristoteles vergleicht den Zustand von Unbeherrschten mit denen von Schlafenden, Wahnsinnigen und Betrunkenen: aufgrund ihres körperlichen Zustandes sind sie nicht in der Lage, ihr dispositionell vorhandenes Wissen auch anzuwenden.

Schließlich gibt Aristoteles in Schritt 4<sup>72</sup> eine naturphilosophische Erklärung akratishen Handelns sowie des Wissens des Akratikers. Hierzu greift er auf die Figur des praktischen Syllogismus zurück: Während in Fällen theoretischer Erkenntnis jemand die Konklusion aus zwei Prämissen notwendigerweise bejahen muss, so muss jemand bei einem praktischen Syllogismus sofort handeln. Ein praktischer Syllogismus besteht aus einem allgemeinen Obersatz, einem partikulären Untersatz, und einer Konklusion, welche mit der Handlung selbst identisch ist:

„Wenn man alles Süße genießen soll, dieses hier als ein bestimmtes Einzelding aber süß ist, dann muss notwendigerweise derjenige, der das Vermögen hat und nicht gehindert ist, dies zugleich auch tun.“ (NE 1147a 29–31)

Der Akratiker verfüge nun über zwei verschiedene allgemeine Prämissen bzw. Obersätze, die auf eine bestimmte Situation angewendet zu entgegengesetzten Handlungen führen würden:

---

<sup>69</sup> Vgl. NE 1147a 1–9.

<sup>70</sup> Vgl. NE 1147a 10–24.

<sup>71</sup> Bei mittelalterlichen Kommentatoren ist dies der Unterschied zwischen *habitus solutus* und *habitus ligatus*. Vgl. Price 2006, 238.

<sup>72</sup> Vgl. NE 1147a 24–b9. Oftmals werden die vorhergehenden drei Schritte als dialektische Lösung im Unterschied zur abschließenden naturphilosophischen (*physikos*) Lösung bezeichnet. Vgl. Robinson 2010, 196–198.

„Wenn also eine allgemeine Meinung in der Seele vorhanden ist, die das Genießen verbietet, und eine andere, dass alles Süße angenehm ist, dies hier aber süß ist (und diese Meinung in Betätigung ist), und wenn nun gerade eine Begierde in der Seele vorhanden ist, dann sagt die eine [Meinung], man solle dies fliehen, die Begierde aber treibt an; denn sie kann jeden Körperteil bewegen.“ (NE 1147a 31–35)

Lesart A versteht diese Ausführungen so:<sup>73</sup> Der Akratiker verfüge über eine vernünftige Prämisse, welche die akratische Handlung in dieser Situation untersagen würde („Ich soll alles Süße meiden“). Doch aufgrund einer übermäßigen Begierde, z.B. hervorgerufen durch einen aktuell wahrgenommenen süßen Gegenstand, wendet der Akratiker die vernünftige allgemeine Prämisse nicht an, sondern bildet einen alternativen akratischen Syllogismus mit einer alternativen allgemeinen Prämisse („Man soll alles Süße kosten“), welche in Gemeinschaft mit der spezifischen partikulären Prämisse („Dieses Einzelding ist süß“) zur akratischen Handlung („Ich koste dieses Einzelding“) führt. Schematisch könnte dies folgendermaßen dargestellt werden:

	<b>Syllogismus der Vernunft</b>	<b>Syllogismus der Begierde</b>
Obersatz	Man soll alles Süße meiden	Man soll alles Süße kosten
Untersatz	<del>Dieses Einzelding ist süß</del>	Dieses Einzelding ist süß
Konklusion (Handlung)	<del>Ich meide dieses Einzelding</del>	Ich koste dieses Einzelding

Der Akratiker weiß, was er am besten tun sollte, indem er über die vernünftige allgemeine Prämisse verfügt. Er verfügt jedoch nicht über das Wissen in Form einer zugehörigen partikulären Prämisse. Deshalb gelangt der richtige praktische Syllogismus nicht zum Abschluss und die Person handelt nicht so, wie sie eigentlich handeln sollte. Das Wissen um die partikuläre Prämisse wird dabei durch eine übermäßige Begierde unterdrückt bzw. dessen Anwendung in der konkreten Situation verhindert. Stattdessen handelt der Akratiker nach einem alternativen Syllogismus, einem Syllogismus der Begierde. So kann Aristoteles sowohl der alltäglichen Meinung, Akrasia komme vor und bestehe im Handeln wider besseren Wissens, und der sokratischen Meinung, Wissen impliziere das Handeln in Übereinstimmung mit diesem Wissen, gerecht werden. Denn nicht das eigentliche Wissen, die allgemeine Prämisse der Vernunft,<sup>74</sup> wird durch den Affekt ‚hin und hergezerrt‘ sondern nur das partikuläre Wissen. Der Akratiker handelt so aus einer spezifischen Unwissenheit, die durch einen Affekt ausgelöst wird, aber gegen sein besseres Wissen in Form der allgemeinen Prämisse, die dispositionell vorliegt oder auch aktuell vom Akteur erwägt wird. Entscheidend ist, dass die zweite, partikuläre Prämisse dem Akteur nicht aktuell zugänglich ist. Denn ansonsten würde die Person gemäß dieser Auffassung des praktischen Syllogismus sofort handeln. Das kognitive Defizit erklärt mithin, weshalb die Person nicht das tut, was sie für das beste hält, sondern etwas anderes.

<sup>73</sup> Vgl. Wolf 2007, 175ff.

<sup>74</sup> Aristoteles gibt nicht an, wie Ober- und Untersatz des Syllogismus der Vernunft genau lauten. Vgl. Owens 1985.

## **b) Probleme mit Lesart A<sup>75</sup>**

1) Wenn das Fehlen der partikulären Prämisse des Untersatzes des Syllogismus der Vernunft verhindert, dass die Person richtig handelt: wie lautet dann diese Prämisse? Wenn sie identisch wäre mit der Prämisse des akratischen Syllogismus (wie in der obigen Darstellung), dann ist unklar, wie sie in einem Fall fehlen, im anderen jedoch aktiv sein könnte.<sup>76</sup> Das heißt: wenn die Person einerseits merkt, dass ein Gegenstand süß ist (Dieses Einzelding ist süß), dann kann der Untersatz des Syllogismus der Vernunft nicht ebenso lauten, da diese ja gemäß dieser Lesart gerade nicht aktuell gegeben ist. Anders formuliert: Wenn der Akratiker deshalb nicht beherrscht handelt, weil er nicht merkt, dass ein Gegenstand süß ist, wie kann er ihn dann aber zugleich kosten, *weil* er süß ist. Verschiedene Vorschläge wurden gemacht, um dieses Problem innerhalb der Grenzen der Lesart A zu lösen:<sup>77</sup>

a) Der Untersatz ist der Person bewusst, aber sie verknüpft ihn nicht mit dem vernünftigen Obersatz, sondern mit dem akratischen. Oder der Untersatz könnte zwar inhaltlich bewusst sein, wird aber nicht mit den übergreifenden Zielen des Akteurs in Verbindung gebracht, wodurch der Widerspruch dem Akteur verborgen bleibt.<sup>78</sup> Problematisch an diesen Versuchen ist unter anderem die Frage, wie der Person ein solch offensichtlicher logischer Schnitzer verborgen bleiben bzw. sie einen solchen überhaupt begehen kann. Hier wird dann oft auf das – philosophisch nicht minder erklärungsbedürftige – Phänomen der Selbsttäuschung verwiesen, das solche Zustände erlaube.

b) Der entsprechende Untersatz besteht aus mehreren Teilen oder Einzelsätzen.<sup>79</sup> So könnte dem Akratiker bloß ein Teil bewusst sein, der für den akratischen Syllogismus ausreicht, und der für die Verbindung mit dem vernünftigen Obersatz relevante Teil nicht. Problematisch daran ist, dass auch hier die Frage bleibt, wie der Person offensichtliche Details der Situation nicht aktuell bewusst sein können, und wie diese Teile innerhalb des praktischen Syllogismus genau lauten, ohne unplausibel zu sein und dennoch das gewünschte Ergebnis zu liefern.

2) Wenn der richtige, vernünftige praktische Syllogismus nicht zum Abschluss gelangt, wieso meint dann Aristoteles, dass der Akratiker gegen seinen Vorsatz (*prohairesis*) handle?<sup>80</sup> Ein Vorsatz wird von Aristoteles folgendermaßen bestimmt

---

<sup>75</sup> Zu der Liste an Problemen vgl. Müller 2009, 118ff.; Bostock 2000, 132ff.

<sup>76</sup> Vgl. Müller 2009, 118; Gosling 1990, 28.

<sup>77</sup> Vgl. Müller 2009, 119f.

<sup>78</sup> Nach Müller 2009, 119f. sind dies zwei verschiedene Reparaturversuche. Ich sehe sie als im Prinzip gleich an.

<sup>79</sup> Vgl. Schritt 2 der aristotelischen Lösung.

<sup>80</sup> Übersetzungsvarianten für ‚*prohairesis*‘: Entscheidung, Wille, Willensentscheidung, Entschließung, Willenswahl, Vorzugswahl; engl. *choice*. Vgl. Wolf 2007, 266. Zum Begriff vgl. auch Bostock 2000, 103–122; Rapp 2010; Liske 2008. „Dass nun die Unbeherrschttheit kein Laster ist, ist offensichtlich (in gewissem Sinn allerdings ist sie es

„Da nun Gegenstand des Vorsatzes etwas Erstrebtes und Überlegtes unter denjenigen Dingen ist, die in unserer Macht stehen, wird auch der Vorsatz ein mit Überlegung verbundenes Streben (*orexis bouleutikē*) nach den Dingen sein, die in unserer Macht (*ta eph' hēmin*) stehen. Nachdem wir nämlich als Ergebnis der Überlegung eine Entscheidung erreicht haben, streben wir der Überlegung entsprechend.“ (NE 1113 a10–13)

Wie gelangt der Akratiker gemäß der Lesart A jedoch zu seinem Vorsatz, wenn er die vernünftige Überlegung in Form des praktischen Syllogismus nicht vollendet? Akrasia in Form der Voreiligkeit, also bei der der Handelnde überhaupt nicht überlegt, könnte so zwar verständlich sein, jedoch nicht die schwache Form der Akrasia.

a) Ein Vorsatz könnte auch als allgemeine Maxime verstanden werden, im Sinne eines Obersatzes eines praktischen Syllogismus. Dieser kann ja nach Lesart A der Person durchaus auch aktuell bewusst sein. Doch wie passt dies zu Äußerungen von Aristoteles, dass ein Vorsatz ein mit Überlegung verbundenes Streben ist?

b) Es wurde vorgeschlagen, dass die akratistische Person doch den vernünftigen Syllogismus vollendet, d.h. zur entsprechenden Konklusion gelangt und dennoch entgegen dieser Konklusion handelt.<sup>81</sup> Dies verlangt jedoch, dass man erstens Konklusionen von praktischen Syllogismen entgegen direkter textlicher Evidenzen<sup>82</sup> nicht als Handlungen, sondern als propositional verfasste Handlungsvorsätze bzw. -anweisungen auffasst, und zweitens, dass man ‚teleutaia protasis‘ in der Zusammenfassung von Aristoteles am Ende des Wissenskapitels nicht mit ‚letzter Prämisse‘ sondern mit ‚letzter Proposition‘, welche auch eine Konklusion sein könnte, übersetzt:

„Da die letzte Proposition [teleutaia protasis] eine Meinung über einen Wahrnehmungsgegenstand enthält und diese das Handeln bestimmt, besitzt der Unbeherrschte, da er sich im Affekt befindet, diese entweder gar nicht, oder er besitzt sie so, dass das Besitzen kein Wissen ist, sondern ein Sprechen wie das eines Betrunknen, der Verse des Empedokles aufsagt.“ (NE 1147 b9)

Beide Punkte sind sehr umstritten<sup>83</sup> und diese alternative Lesart sprengt letztlich den interpretatorischen Rahmen der Lesart A: Wenn das Wissen um die Konklusion vollständig aktuell einer Person gegeben ist, wie kann dann noch die sokratische handlungstheoretische Grundprämisse, der gemäß Wissen hinreichend ist für Handeln, aufrecht erhalten werden, wenn die Person in diesem Fall trotzdem entgegen diesem Wissen handelt?

3) Der Konflikt des Akratikers wird in dieser Lesart nur als diachroner Konflikt und als rein kognitives Defizit fassbar. Diachron deshalb, da synchrone, klarsichtige Willensschwäche nach diesem Modell nicht möglich ist. Rein kognitiv ist der Konflikt deshalb, da die unvernünftige

---

vielleicht). Denn das eine [das unbeherrschte Handeln] ist gegen den Vorsatz, während das andere [die Handlung des Lasters] dem Vorsatz entspricht.“ (NE 1151 a5–7) Vgl. auch NE 1148 a5–20; NE 1111 b10–17.

<sup>81</sup> Vgl. Müller 2009, 122f.

<sup>82</sup> Aristoteles 1985, 6.701 a12–13

<sup>83</sup> Bereits Walter Burley hatte im Mittelalter vorgeschlagen, hier mit ‚Schlussfolgerung‘ statt ‚letzter Prämisse‘ zu übersetzen. Vgl. Müller 2009, 691; Charles 2009, 42; Saarinen 1994, 131–146.

Begierde lediglich dafür sorgt, dass der Akratiker in der bestimmten Situation erst gar nicht sein Wissen, dass eine bestimmte Handlung die beste ist, aktualisieren kann. Läge dieses Wissen vor, so würde der Akratiker entsprechend handeln.<sup>84</sup> Oder anders formuliert: aktuell weiß die handelnde Person nicht, dass sie anders handeln sollte, als sie es tut. Wüsste sie es, so würde sie entsprechend handeln. Dies widerspricht außerdem anderen Stellen über Akrasia außerhalb des Buches NE VII, in denen Aristoteles sich offen zeigt für eine synchrone Fassung des Konflikts.

4) Weshalb ist der Akratiker für seine willensschwachen Handlungen verantwortlich, wenn Unwissenheit über partikuläre Handlungsumstände nach Aristoteles ein Handeln unwillentlich machen und gemäß dieser Lesart eine solche Unwissenheit vorliegt?<sup>85</sup>

a) Nach Aristoteles tragen Personen auch Verantwortung für ihre Dispositionen, aus denen Handlungen hervorgehen. Der Unbeherrschte ist verantwortlich für seine affektiven Dispositionen sowie seinen Umgang mit diesen, daher ist er indirekt verantwortlich für seine akkratischen Handlungen, auch wenn er in der jeweiligen Situation nicht anders als akkratisch hätte handeln können.

b) Aristoteles unterscheidet zwischen Handlungen *aus* Unwissenheit und Handlungen *in* Unwissenheit. Letztere sind der Person zurechenbar, wenn die Person verantwortlich ist für ihre Unwissenheit, wie z.B. im Falle von Betrunkenen. Obwohl der Akratiker in Unwissenheit handelt, da er gemäß der Lesart A wesentliche Merkmale der Situation bzw. der Handlungsumstände nicht kennt, ist er für diese Unwissenheit verantwortlich, da sie durch einen übermäßigen Affekt hervorgerufen wird, für den er verantwortlich ist.<sup>86</sup>

### **c) Lesart B**

Lesart B<sup>87</sup> interpretiert die aristotelische Position im Unterschied zu Lesart A als grundsätzlich verschieden von der sokratisch-platonischen Position.<sup>88</sup> Aristoteles teilt nicht den sokratischen handlungstheoretischen Grundsatz – Wissen von X beinhaltet Handeln in Übereinstimmung mit X – und kann daher die Bestimmung von Willensschwäche – Akrasie ist ein Handeln wider bes-

---

<sup>84</sup> Vgl. Ross 1923, 232.

<sup>85</sup> „Da gegen das Wollen (*akousion*) das ist, was durch Zwang und aufgrund von Unwissenheit geschieht, dürfte als das Gewollte (*hekousion*) dasjenige gelten, dessen Ursprung (*arche*) im Handelnden selbst liegt, wobei er die einzelnen Bedingungen (*ta kath' hekesta*) kennt, unter denen die Handlung stattfindet.“ (NE 1111 a21–24) Vgl. Müller 2009, 141–152. Für diese Lesart vgl. Charles 1984; Charles 2009; Aristoteles 2002, 385–387.

<sup>86</sup> Allgemein zum Begriff der Verantwortung bei Aristoteles vgl. Rapp 2010.

<sup>87</sup> Für diese Lesart vgl. Charles 1984; Broadie 1991; Dahl 1984; Corcilius 2008b; Aristoteles 1995.

<sup>88</sup> Zur platonischen Position im Unterschied zur sokratischen vgl. Platon 2008, 434c–445e; Rebutisch 2009; Morris 2006; Penner 1992. Hier ist nicht der Raum, um auf die Unterschiede zwischen diesen beiden Positionen einzugehen. Es sei nur erwähnt, dass Platon über seine Theorie der Seelenteile über Möglichkeiten verfügt, anders bzw. für einige adäquater mit dem Phänomen der Willensschwäche umzugehen als Sokrates.

seres Wissen – vollauf gelten lassen.<sup>89</sup> Damit ist prinzipiell die Möglichkeit gegeben, synchrone motivationale Willensschwäche theoretisch zu akzeptieren. Bei der Einführung des Wissensproblems<sup>90</sup> stellt Aristoteles die sokratische Position vor, seine eigene Problemformulierung ist jedoch von der sokratischen Problemstellung verschieden, da er gerade die sokratische handlungstheoretische Grundprämisse nicht teilt.<sup>91</sup> Dies bedeutet im fehlerhaften Umkehrschluss natürlich nicht, dass es keine Formen von Willensschwäche gäbe, bei denen fehlendes Wissen bzw. aktuelle Unwissen Ursachen willensschwacher Handlungen sind. Dies heißt nur, dass auch bei aktuell vorliegendem Wissen eine diesem Wissen entsprechende Handlung nicht notwendig folgt bzw. einhergeht. Was ist aber stattdessen das handlungstheoretische Grundmodell des Aristoteles gemäß dieser Lesart B? Für Aristoteles ist Wissen alleine keine hinreichende Bedingung für Handlungen, sondern nur Wissen in Kombination mit einer Strebung (orexis).<sup>92</sup>

„Aristoteles entwickelt seine Psychologie menschlichen Handelns in *De anima* III 9–11 und *De motu animalium* 6–8 charakteristischerweise innerhalb einer allgemeinen Theorie der Ortsbewegung von Lebewesen. Die zentrale Frage lautet: „Was ist das dem Orte nach das Lebewesen bewegende Prinzip?“<sup>93</sup> Die aristotelische Antwort unterscheidet zwischen einem äußeren Prinzip der Bewegung, dem Strebensobjekt (orekton), das als Finalursache fungiert, und zwei inneren Prinzipien in Gestalt von Streben (orexis) und Vernunft (nous). Die beiden letztgenannten sind ihrerseits generische Begriffe, die noch weitere Spezies umfassen: Unter den hier „unterterminologisch“ gebrauchten Terminus nous fallen Überlegung (dianoia), Vorstellung (phantasia) und sinnliche Wahrnehmung (aisthesis); die orexis umfasst den Wunsch (boulesis), der dem rationalen Seelenvermögen zugerechnet wird, sowie Begierde (epithymia) und Eifer bzw. Zorn (thymos), die dem irrationalen Seelenteil angehören.“ (Müller 2009, 130; gr. Buchstaben durch lat./dt. ersetzt v. R.H.)

Für Aristoteles sind demnach immer zwei Elemente nötig, um eine Handlung effizient-kausal zu bewirken: ein kognitives Element in Form einer Überlegung, Vorstellung oder sinnlichen Wahrnehmung, und ein strebendes Element in Form eines Wunsches, einer Emotion (thymos) oder einer Begierde. Daher kann Wissen (episteme) alleine keine Handlungen verursachen, sondern es bedarf dazu des Zusammenspiels zwischen diesen beiden Elementen.<sup>94</sup> Auf das Phänomen der Akrasia angewendet ergibt sich die Konsequenz, dass es prinzipiell möglich ist,

<sup>89</sup> Zur Formulierung des sokratischen Dilemmas s.o. und vgl. Corcilius 2008b, 145.

<sup>90</sup> S.o. und vgl. NE 1145 b21–31.

<sup>91</sup> „Anstatt das Offensichtliche zu bestreiten, so die Kritik des Aristoteles, hätte Sokrates untersuchen sollen, welche Weise der Unkenntnis es ist [...], aufgrund derer akratisches Verhalten zustande kommt, *wenn* es aufgrund von Unkenntnis zustande kommt. Dabei wird die These, *dass* akratisches Verhalten aufgrund von Unkenntnis zustande kommt, von Aristoteles nicht unterstützt, sondern er argumentiert hier nur auf dem Boden der Sokratischen Position: Selbst wenn es sich um Unkenntnis handeln sollte, so sein Argument, wäre erforderlich gewesen, genauer zu bestimmen, welches die Weise dieser Unkenntnis ist.“ (Corcilius 2008b, 145)

<sup>92</sup> Vgl. Müller 2009, 130ff.

<sup>93</sup> Aristoteles 1995, III.9 432 b9.

<sup>94</sup> „Das Denken als solches bewegt jedoch nichts, sondern nur dasjenige Denken, das auf einen Zweck bezogen, das heißt praktisch ist.“ (NE 1139 a36)

dass eine Person über das Wissen verfügt, was sie in einer Situation tun sollte, und es dennoch nicht tut, da sie das Ziel der Handlung nicht oder nicht ausreichend erstrebt.<sup>95</sup>

Nun stellen sich dieser Lesart einige interpretatorische Hürden entgegen:

1) Die Vergleiche des Akratikers mit Betrunkenen, Wahnsinnigen und Schlafenden<sup>96</sup> legen eine gewisse Art der Unwissenheit des Akratikers nahe, die für Lesart A spricht. Lesart B kann diese Analogien jedoch auch anders verstehen:<sup>97</sup> Was verglichen wird, ist nicht die jeweilige Unwissenheit, sondern das spezifische Verhältnis, in dem die Person zu ihrem vorhandenen Wissen steht.<sup>98</sup> Hierzu passen auch die weiteren Hinweise auf Schauspieler und Lernende: erstere sprechen in einer uneigentlichen Weise, bei letzteren ist das Wissen noch nicht „mit dem Menschen verwachsen“ (NE 1147 a23). Schauspieler streben bloß innerhalb ihrer fiktionalen Rolle nach den Gütern, von denen sie wissen, und Lernende können zwar den Gehalt ihres Wissens korrekt wiedergeben, ohne dieses Wissen jedoch auch richtig anzuwenden. Also weder die Gehalte, noch die rein epistemische Relation des Akratikers zu seinem Wissen wird verglichen, sondern lediglich seine affektiv-strebende Relation zu seinen Wissensgehalten, die ihm mehr oder weniger gegeben sind.

Dieser Interpretationsmöglichkeit steht vor allem entgegen, dass Aristoteles am Ende seiner naturphilosophischen Erläuterung des Herganges der Unbeherrschtheit explizit von der Aufhebung der Unwissenheit spricht:

„Wie aber die Unwissenheit aufgelöst und der Unbeherrschte wieder wissend wird, darüber gibt es für diesen affektiven Zustand keine eigene Erklärung, sondern es gilt hier dieselbe Erklärung wie für den Betrunkenen und Schlafenden; dazu muss man die Naturforscher hören.“ (NE 1147 b6–9)

Vielleicht sollte man diese Bemerkung ebenfalls im Lichte der zuvor angestellten Analogien zu den Betrunkenen und Schlafenden lesen, also nicht als Unwissenheit bestimmter Gehalte, sondern als spezifische Relation, nämlich einer mangelnden Strebung nach diesen Inhalten. Willensschwache Charaktere haben demnach temporär (Betrunkener, Schlafender) oder dauerhaft (Schauspieler, Lernender) eine mangelhafte rationale Strebung, ein mangelhaftes Ausmaß an Vernünftigkeit als rationaler Motivation. Dazu passt die Bemerkung, dass willensschwache Charaktere wie Leute wären, die von zu wenig Wein betrunken werden: zu

---

<sup>95</sup> Man könnte Lesart A auch als belief-approach und Lesart B als desire-approach bezeichnen, je nachdem, wo der Mangel des Akratikers festgemacht wird. Diese Unterscheidung berücksichtigt jedoch keine Mittelpositionen, welche gerade auf das enge Zusammenspiel von beliefs und desires abheben.

<sup>96</sup> Vgl. NE 1147 a10–24.

<sup>97</sup> Vgl. Corcilius 2008b, 152–157.

<sup>98</sup> „Das Problem ist nicht, dass sie etwas nicht wissen, sondern dass sie als Personen in einer für die Anwendung des Wissens falschen *Relation* zum Gehalt ihres Wissens stehen.“ (Corcilius 2008b, 154)

wenig natürlich gemessen an der jeweiligen Situation und den als normal gesetzten Standard einer Gesellschaft.<sup>99</sup> Dies würde Willensschwäche als teilweises soziales Phänomen deuten.<sup>100</sup>

2) Lesart B liest oftmals entgegen einiger Bedenken ‚teleutaia protasis‘ als ‚letzte Proposition‘, nicht als ‚letzte Prämisse‘ eines Syllogismus, und identifiziert diese mit der Konklusion.<sup>101</sup> Wie in Eckpunkt E6 oben erwähnt unterscheidet Aristoteles zumindest zwischen zwei Grundformen von Akrasia, einer voreiligen (propetaia) und einer schwachen (asthenaia). Im Falle von voreiliger Akrasia überlegt die Person nicht, und besitzt daher die letzte Proposition d.h. Konklusion des Syllogismus der Vernunft nicht. Dieser Fall wäre demnach recht analog zu Lesart A. Im anderen Falle jedoch, bei schwacher Akrasia, überlegt die Person und gelangt zur richtigen Konklusion, jedoch besitzt sie diese nur wie ein Betrunkener oder Schlafender, d.h. sie hat zwar Wissen von dieser, jedoch gelangt dieses Wissen nicht zur Ausführung da es nicht oder nicht ausreichend erstrebt wird, weil ein starker Affekt zugleich vorhanden ist. Nach dieser Lesart gelangt also zumindest ein Typus des Akratikers zur richtigen Konklusion, wird jedoch durch einen Affekt an der Ausführung der Konklusion des praktischen Syllogismus gehindert und handelt stattdessen nach einem alternativen Syllogismus der Begierde.<sup>102</sup> Damit wäre synchrone, klarsichtige Willensschwäche im Unterschied zu Lesart A möglich.

3) Lesart B interpretiert Rolle und Funktion des praktischen Syllogismus anders als Lesart A. Wenn der eine Typus des Akratikers zur richtigen Konklusion gelangt, jedoch nicht nach dieser handelt, dann kann die Konklusion eines praktischen Syllogismus nach dieser Auffassung nicht mit einer Handlung identisch sein. Nun legen jedoch einige Stellen, in denen Aristoteles von der Figur des praktischen Syllogismus Gebrauch macht, gerade eine solche Identifikation nahe:

„Wie aber (ist es zu erklären, dass) man, wenn man denkt, bisweilen handelt, bisweilen dagegen nicht und sich bewegt, sich bisweilen aber nicht bewegt? Eine ähnliche Situation scheint auch gegeben zu sein wenn man im Hinblick auf die Dinge, die keine Bewegung haben, Überlegungen anstellt und logische Schlüsse zieht. Aber hier ist das Endergebnis ein wissenschaftlicher Satz – denn wenn man die zwei Prämissen denkt, dann denkt man auch die Schlussfolgerung und setzt sie zusammen –, dort aber ist die Schlussfolgerung aus den zwei Prämissen eine Handlung; zum Beispiel: Wenn man die Überlegung anstellt, dass jeder Mensch gehen muss, man selbst aber ein Mensch ist, geht man sofort, wenn (man) aber (die Überlegung anstellt,) dass zum gegenwärtigen Zeitpunkt kein Mensch gehen darf, man selbst aber ein Mensch ist, verharrt man sofort im Zustand der Ruhe; und in diesen beiden Fällen handelt man, wenn nicht irgend etwas (daran) hindert oder (dazu) zwingt.“ (Aristoteles 1985, 7.701 a7–17)

---

<sup>99</sup> Vgl. NE 1151a 1–5.

<sup>100</sup> Vgl. Kap. VI.4.

<sup>101</sup> Vgl. Charles 2009, 52–64. Vgl. zuvor die Probleme mit Lesart A unter 2) b).

<sup>102</sup> „In one case of *akrasia*, one proposition emerges from the two premisses, but the agent is dragged away by the activity of desire. This *akratēs* uses the relevant minor premiss ‘This is sweet’ to get to the conclusion ‘Flee this!’ (or ‘I ought to flee this!’). This *akratēs* will be aware of the good conclusion he says, even though he does not act on it.” (Charles 2009, 59)

Es gibt nun verschiedene Möglichkeiten, auf dieses Problem zu reagieren:

a) Man könnte meinen, dass ‚sofort‘ in ‚geht man sofort‘ oder ‚verharrt man sofort‘ eine zeitliche Lücke zwischen propositionaler Konklusion und darauffolgender Handlung zulässt und damit gerade die Nichtidentität der Konklusion und der Handlung voraussetze. Man könnte außerdem einwenden, dass der Affekt des Akratikers eines jener hindernden Umstände ist (wenn nicht irgend etwas (daran) hindert), der die Handlung ausbleiben lässt.<sup>103</sup>

b) Man bewahrt die Identität von Konklusion und Handlung indem man die Figur des praktischen Syllogismus nur als *Analogie* für die effizient-kausale Notwendigkeit, mit der aus bestimmten Bedingungen Handlungen folgen, versteht, statt als Illustration praktischer Deliberation, also als Darstellung, wie tatsächlich aus bestimmten Propositionen notwendig Handlungen hervorgehen.<sup>104</sup> Diese Interpretation hat vor allem Klaus Corcilius vertreten. Für diese Lesart spricht vor allem, dass die Figur des praktischen Syllogismus Kernstück der aristotelischen Theorie der animalischen Ortsbewegung ist. Da diese Theorie auf alle möglichen Lebewesen zutreffen soll, die sich bewegen, ist es sehr unwahrscheinlich, dass das Kernstück dieser Theorie nur für den Spezialfall menschlicher Handlungen, bei denen Propositionen eine Rolle spielen, zutrifft, da nur wenige Tiere wahrscheinlich über Propositionen verfügen und sich dennoch bewegen. Dies fügt sich ein in das grundlegende Bild der aristotelischen Handlungstheorie als Spezialfall der allgemeinen Theorie animalischer Ortsbewegung. Der praktische Syllogismus ist eine Analogie: in Fällen theoretischer Schlüsse folgt die Konklusion *notwendig* aus den Prämissen, so wie bei Vorliegen bestimmter Bedingungen *notwendig* bestimmte Bewegungen folgen.<sup>105</sup> Was sind nun diese Bedingungen? Ohne auf die Details einzugehen, kommt Bewegung nur durch das Zusammenspiel eines kognitiven Elementes und einer Strebung zustande.<sup>106</sup> Bei Akrasia kommt es zu einem Konflikt zwischen verschiedenen Strebungen: ein Wunsch (*boulesis*) nach einem durch die Vernunft vorgestelltem Gut steht in Widerspruch zu einer auf einen Wahrnehmungsgegenstand bezogenen Begierde (*epithymia*) oder Emotion (*thymos*). Was letztlich dafür sorgt, dass sich beim Akratiker die Begierde gegenüber dem

---

<sup>103</sup> Vgl. Charles 2009, 53–55 als Interpretation von NE 1147 a29–31: „‘At the same time‘ (*hama*) refers to the same point as indicated by the phrase ‘immediately act‘ (*enthu prattein*): the point reached after one proposition (or opinion) emerges from the premisses. So understood, any interfering or preventing factors must strike between his drawing the conclusion and acting. This is why one gets to the conclusion and also straightaway acts on it unless prevented from doing so.“ (Charles 2009, 53)

<sup>104</sup> Vgl. Corcilius 2008d.

<sup>105</sup> „Aristoteles möchte mit dem PS [Praktischen Syllogismus] keine inferentiellen Vorgänge, sondern einen kausalen Sachverhalt illustrieren, der für seine Theorie der animalischen Ortsbewegung von zentraler Bedeutung ist. Die Figur des Syllogismus dient ihm als Modell, das verstehen hilft, wie sich Vorkommnisse von animalischen Ortsbewegungen aus gegebenen Antezedenzen mit Notwendigkeit ergeben.“ (Corcilius 2008a, 314)

<sup>106</sup> „Der Ausgangspunkt (*archê*) der animalischen Ortsbewegung [...] besteht in dem Komplex aus Kognition (Wahrnehmung/Vorstellung/Vernunft) mit einer Strebung, die sich auf den durch die Kognition präsentierten Gehalt richtet. Weder Kognition noch Strebung sind jeweils allein dafür hinreichend.“ (Corcilius 2008a, 285)

vernünftigen Wunsch durchsetzt, ist nicht ganz klar. Doch bezeichnet Akrasia genau diese Disposition, in Fällen solcher Konflikte der Begierde statt dem vernünftigen Wunsch zu folgen. Grundlegend lässt die aristotelische Handlungspsychologie zwei Möglichkeiten offen: Erstens: die willensschwache Person hat eine zu schwache rationale Streben (boulesis) – dies passt auf die schwache Form der Akrasia (astheneia). Zweitens: die Begierde der willensschwachen Person ist zu stark – dies passt zur Akrasia als Impulsivität (propreteia). Es bleibt jedoch, dass das Denken der jeweiligen Gehalte in syllogistischer oder nicht-deduktiver Form alleine nach diesem Modell niemals hinreichend sein kann um eine Handlung oder allgemein Bewegung hervorzurufen. Noch allgemeiner fügt sich diese Interpretation recht gut in die Kritik von Aristoteles an der sokratischen Identifikation von Tugend und Wissen ein:

„Deswegen sagen einige, alle Tugenden seien Arten der Klugheit, und deswegen war Sokrates in einer Hinsicht auf dem richtigen Weg, während er sich in anderer Hinsicht im Irrtum befand. Darin nämlich, dass er alle Tugenden für Arten der Klugheit hielt, irrte er, während er mit der Behauptung, dass sie nicht ohne Klugheit vorkommen würden, Recht hatte. [...] Sokrates meinte also, die Tugenden *seien* Überlegungen (sie seien nämlich, so glaubte er, alle Arten des Wissens). Wir hingegen glauben, dass sie *mit* Überlegung *verbunden* sind. Aus dem Gesagten ist also klar, dass man weder im eigentlichen Sinn gut sein kann ohne die Klugheit noch klug ohne die Tugend des Charakters.“ (NE 1144 b18–22; b28–32)

Die Interpretation der aristotelischen Auffassung des Verhältnisses von dianoetischen (insbesondere die Klugheit) und ethischen Tugenden ist natürlich Gegenstand vieler Kontroversen. Sollte dieses Verhältnis jedoch so interpretiert werden, dass Wissen nicht hinreichend ist für zumindest die ethischen Tugenden, dann würde dies dafür sprechen, auch im Falle des Akratikers davon auszugehen, dass er über das Wissen verfügen könnte, und dennoch nicht entsprechend handelt. Lesart B würde sich demnach besser in das naturphilosophische Grundmodell der animalischen Ortsbewegung, die darauf aufbauende aristotelische Handlungstheorie und den Gesamtkontext der ethischen Betrachtung der charakterlichen Tugenden einfügen als Lesart A. Klar ist außerdem, dass die Definition von ‚praktisches Wissen‘ für die Diskussion keinen Nutzen hat: Wer ‚praktisches Wissen‘ so *definiert*, dass jemand, der über solch ein Wissen verfügt auch entsprechend handelt, der kommt aus definatorischen Gründen zu dem Schluss, dass die Akratikerin nicht über das nötige praktische Wissen verfügt. Dies ist aber eine Scheinlösung, da sie die gehaltvolle Diskussion darüber, was denn genau dieses praktische Wissen ist und woraus es besteht, verdeckt.

## II.2 Willensschwäche bei Donald Davidson

Die philosophische Diskussion bei Aristoteles wurde innerhalb einer moralphilosophischen Perspektive geführt. Das Problem, das Willensschwäche aufwarf, war dementsprechend ein moralphilosophisches Problem und musste als ein solches gelöst werden. Demgegenüber wurde Willensschwäche innerhalb der modernen philosophischen Diskussion vorwiegend aus einer

handlungstheoretischen Perspektive diskutiert. Diese Entwicklung ist einerseits mit dem Aufkommen moderner Handlungstheorie allgemein verknüpft, andererseits mit der einflussreichen Behandlung von Willensschwäche durch Donald Davidson, dessen Spielart einer kausalen Handlungstheorie lange Zeit die Diskussion geprägt hatte. Ich werde hier zuerst die Idee der kausalen Handlungstheorie allgemein und besonders in der Form von Davidson vorstellen (1). In (2) diskutiere ich Davidsons Behandlung von Willensschwäche in seinem Aufsatz *How is weakness of the will possible?* (1969a), in dem sich zeigt, dass und wie Willensschwäche als Phänomen untrennbar mit kausalen Handlungstheorien verknüpft ist. Diese Diskussion wird dann ergänzt mit seinen späteren Ansichten in *Paradoxes of irrationality*, in der er eine Erklärung für willensschwache und allgemein irrationale Handlungen gibt (3). Danach folgt eine Diskussion von Kritik an und Problemen mit Davidsons kausaler Handlungstheorie (4) und seiner Konzeption von Willensschwäche (5).

### II.2.1 Kausale Handlungstheorie

Unter ‚kausaler Handlungstheorie‘ verstehe ich eine Familie von verschiedenen Theorien über den Handlungsbegriff und die Struktur von Handlungserklärungen. Die Bezeichnung ‚kausale Handlungstheorie‘ ist, trotz der Popularität solcher Theorien, recht vage. Zum einen kann mit kausaler Handlungstheorie eine Auffassung bezüglich der Natur von Handlungserklärungen über die Angabe von Gründen gemeint sein, sogenannte Rationalisierungen oder intentionale Erklärungen. *Kausalisten* vertreten die These, dass Rationalisierungen im Grunde Kausal-erklärungen sind, während *Anti-Kausalisten* dies bestreiten.<sup>107</sup> Dies ist nicht zu verwechseln mit der logischen Struktur der Handlungserklärung: Für Vertreter *nomologischer* Handlungstheorien ist die logische Struktur von Rationalisierungen nicht verschieden von anderen wissenschaftlichen Theorien nach dem deduktiv-nomologischen Verfahren, wie es v.a. von C. G. Hempel und Paul Oppenheim erarbeitet wurde.<sup>108</sup> Eine Handlung wird erklärt, wenn sich das Eintreten der Handlung aus Gesetzesaussagen und spezifischen Ausgangsbedingungen ableiten lässt. *Nicht-nomologische* Handlungstheorien bestreiten, dass Rationalisierungen diese Struktur haben, d.h. auf Gesetze Bezug nehmen müssen, um Handlungen zu erklären. Außerdem unterscheiden sich die Theorien anhand der Gründe, die sie für ihre Theorien vorbringen – kausal weil nomologisch; kausal weil nicht-nomologisch etc. –, als auch in ihrem Verständnis von Kausalität.<sup>109</sup>

---

<sup>107</sup> Dies ist historisch gesehen der Kern der Debatte in der Zeit Davidsons.

<sup>108</sup> Vgl. Hempel & Oppenheim 1948.

<sup>109</sup> Dabei liegt es nahe, dass jede kausale Handlungstheorie über ein spezifisches Verständnis des Konzepts der Verursachung verfügen muss. Aber auch intentionale Handlungstheorien sollten sich klar sein, welchen Platz sie dem Konzept der Kausalität in unserem Begriffsnetz einräumen und wie sie dessen Rolle verstehen. Versteht man den Begriff der Kausalität weit, so würde auch Kants ‚Handlungstheorie‘ Handlungserklärungen als kausal verstehen, nämlich als Kausalität aus Freiheit.

In einem Schema ergeben sich vier mögliche grundlegende Positionen:

Rationalisierungen	<i>Logische Struktur</i>	
<i>Interpretation</i>	<b>nomologisch</b>	<b>nicht-nomologisch</b>
<b>kausal</b>	Paul Churchland, C. G. Hempel	Donald Davidson
<b>nicht-kausal</b>	Immanuel Kant	William Dray, Abraham Melden

Zum anderen explizieren kausale Handlungstheorien den Begriff der ‚Handlung‘ (oder ‚intentionale Handlung‘) mit Rückgriff auf eine kausale Relation, die gemäß dieser Theorien zwischen bestimmten (mentalen) Ereignissen oder Zuständen und Körperbewegungen oder allgemeiner Ereignissen in der Welt bestehen:

Eine Handlung liegt<sup>110</sup> dann vor, wenn ein mentales Ereignis X ein Ereignis Y verursacht.<sup>111</sup>

Dabei unterscheiden sich verschiedene Handlungstheorien darin, was sie für X und Y einsetzen und wie sie die kausale Relation verstehen. Intuitiv ist die kausale Analyse in der alltäglichen Erfahrung verankert, da bei (vielen) Handlungen stets etwas passiert bzw. eine Veränderung in der Welt hervorgerufen wird. Sprachlich zeigt sich dieser Umstand darin, dass viele Beschreibungen von Handlungen logisch das Eintreten bestimmter Resultate implizieren. So trifft die Beschreibung einer Handlung wie ‚Daniel schaltet das Licht ein‘ nur dann zu, wenn das Licht eingeschaltet wird. Die Frage, welche die kausale Handlungstheorie beantworten will, ist, wie sich diese Resultate von Handlungen von anderen Resultaten, wie z.B. bloßen Ereignissen, unterscheiden.<sup>112</sup> Das Licht könnte eingeschaltet werden, ohne dass Daniel oder irgendjemand anderer es einschaltet. Allgemein formuliert diesen Umstand Lawrence H. Davis:<sup>113</sup>

„Corresponding to each type of doing A will be a type of event  $E_A$  such that “Someone did an A” (read “Someone did something of type A”) implies “There was an  $E_A$  (i.e., an event of type  $E_A$ ),” but “There was an  $E_A$ ” does not imply “Someone did an A.”” (Davis 1979, 6)

Handlungstheorien, die den Anspruch erheben, den Begriff der Handlung zu explizieren, haben in diesem Zusammenhang ein grundlegendes Problem zu lösen: Die Explikation wird erst dann

<sup>110</sup> Ich verwende den neutralen Ausdruck ‚vorliegen‘, da verschiedene kausale Handlungstheorien sich darin unterscheiden, was sie in dieser Relation jeweils als ‚Handlung‘ identifizieren.

<sup>111</sup> „[...] [C]ausal theories of action view actions in terms of their typical causes: actions are those items of behaviour that are caused by certain antecedent events, characteristically beliefs and desires.” (Moya 1990, 44) „According to familiar *causal* theories of action, actions are events that are caused in a certain way (or ways).“ (Mele 1992, 199) „The basic principle common to all causal theories is that *the agent performs an action only if an appropriate internal state of the agent causes a particular result in a certain way.*” (Davis 2010, 32) Vgl. Bishop 1989, 95ff.; Hornsby 1980; Goldman 1970, 70ff.

<sup>112</sup> „The fundamental question of action theory is what makes a piece of behavior qualify as an action. We can distinguish bringing something about from something's merely happening to us, but this serves not to answer the question but only to rephrase it.” (Bach 1978, 361)

<sup>113</sup> Gegen die Annahme, jede Handlung impliziere notwendig das Stattfinden eines Ereignisses, argumentieren unter anderem Moya 1990 und Bach 1980.

erfolgreich sein, wenn das Explikat keine Begriffe enthält, welche das Explikandum für ihr Verständnis bereits voraussetzen. Also im Falle von ‚Handlung‘ keine Begriffselemente, die wiederum nur durch den Begriff der Handlung verständlich sind. Dies kann als Adäquatheitsbedingung für eine gelungene Explikation des Handlungsbegriffs bezeichnet werden. Davidson formuliert diese Adäquatheitsbedingung an kausale Handlungstheorien folgendermaßen:<sup>114</sup>

„The only hope for the causal analysis is to find states or events which are causal conditions of intentional actions, but which are not themselves actions or events about which the question whether the agent can perform them can intelligibly be raised. The most eligible such states or events are the beliefs and desires of an agent that *rationalize* an action, in the sense that their propositional expressions put the action in a favourable light, provide an account of the reasons the agent had in acting, and allows us to reconstruct the intention with which he acted.” (Davidson 1973, 72)

Die kausalen Handlungstheorien unterscheiden sich darin, *was* sie als Handlung in der Verursachungsbeziehung ( $X \rightarrow Y$ ) identifizieren. Hier lassen sich zumindest drei grundlegende Positionen voneinander unterscheiden:<sup>115</sup>

a) Handlung = X: Handlungen sind eigentlich bestimmte mentale Ereignisse wie Willensakte oder Versuche (tryings).<sup>116</sup>

b) Handlung = Y: Handlungen sind Wirkungen von bestimmten mentalen Ereignissen oder Gründen, die wiederum als mentale Zustände wie Wünsche, Überzeugungen, Willensakte oder Intentionen verstanden werden.<sup>117</sup>

c) Handlung =  $X \rightarrow Y$ : Eine Handlung ist das zusammengesetzte Ereignis bzw. das Vorkommen einer bestimmten Art der Verursachung.<sup>118</sup>

Kausale Theorien können also als Theorien der Explikation des Handlungsbegriffs oder als Theorien über die richtige Art der Handlungserklärung formuliert werden. Oftmals wird eine enge Verbindung zwischen diesen beiden Vorhaben angenommen: eine Handlung ist dann etwas,

---

<sup>114</sup> Vgl. auch Davidson 1987. Diese Adäquatheitsbedingung trifft natürlich auf jede Begriffsexplikation zu.

<sup>115</sup> „By definition, any causal theory holds that the production of a change counts as the performance of an action just in case that change is caused in the right way by the right sort of psychological event. What constitutes ‘right’ here is where the various versions of the theory diverge, but they all agree that an action involves one event causing another. They diverge also on whether an action is to be identified with one, the other, or both of these events.” (Bach 1980, 116)

<sup>116</sup> Jennifer Hornsby: „Every action is an event of *trying* or attempting to act, and every attempt that is an action precedes and causes a contraction<sub>i</sub> of muscles and a movement<sub>i</sub> of the body.” (Hornsby 1980, 33) Mit dem Index *T* markiert Hornsby den intransitiven Gebrauch des entsprechenden Wortes.

<sup>117</sup> Alfred Mele: „For all actions *A*, *A* is an intentional action only if *A*’s agent had a reason for *A*-ing and (his having) that reason was a cause of his *A*-ing.” (Mele 1987a, 32)

<sup>118</sup> John Stuart Mill: „Now what is an action? Not one thing, but a series of two things: the state of mind called a volition, followed by an effect. The volition or intention to produce the effect, is one thing; the effect produced in consequence of the intention is another thing; the two together constitute the action.” (Mill 1974, I, ch. iii, § 5, 55)

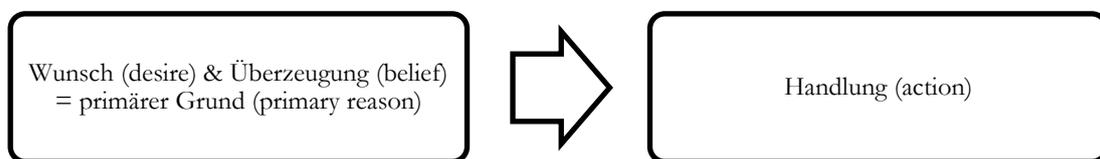
was auf bestimmte Art erklärt werden kann; oder etwas kann auf eine bestimmte Art erklärt werden, wenn es eine Handlung ist.<sup>119</sup>

Davidsons wegweisender Aufsatz *Actions, reasons, and causes* (1963) erschien zu einer Zeit als die philosophische Diskussion bestimmt war von anti-kausalistischen Konzeptionen der Natur und Erklärung von Handlungen.<sup>120</sup> In Davidsons kausaler Handlungstheorie sind Handlungen Ereignisse, genauer Körperbewegungen (Y), welche durch primäre Gründe (X) verursacht werden. Primäre Gründe (primary reasons) bestehen aus Wünschen bzw. Pro-Einstellungen (desires bzw. pro-attitudes)<sup>121</sup> und entsprechenden Überzeugungen (beliefs) einer Person. Davidsons grundlegende These in diesem Aufsatz ist, dass Handlungserklärungen über die Angabe von Gründen der Akteure für ihre Handlungen Kausalerklärungen sind.<sup>122</sup> Davidsons zwei Thesen lauten:

1) „R is a primary reason why an agent performed the action *A* under the description *d* only if R consists of a pro attitude of the agent towards actions with a certain property, and a belief of the agent that *A*, under the description *d*, has that property.” (Davidson 1963, 5)

2) „A primary reason for an action is its cause.” (Davidson 1963, 12)

Mein Wunsch, eine Banane zu essen, in Gemeinschaft mit der Überzeugung, dass eine Banane in der Küche liegt und dass ich sie holen müsste, um sie zu essen, bilden zusammen den Grund, der verursacht, dass ich in die Küche gehe und mir eine Banane hole. Schematisch kann dies folgendermaßen ausgedrückt werden:



<sup>119</sup> Vgl. diese Verquickung bei Paul Churchland und Frederick Stoutland: „An event-description of the general form “*X A*-ed” is a description of a full-blooded action if and only if there exists an explanans deductively entailing “*X A*-ed” [...]” (Churchland 1970, 229) „A causal theory sees these two claims as necessarily connected: an intentional act just is an act that has a certain kind of explanation, namely, one that was done for a reason, and an act is done for a reason only if it is caused in a certain way.” (Stoutland 1980, 351)

<sup>120</sup> Zur historischen Entwicklung vgl. Keil 2000, 13–20. Wie für philosophische Diskussionen üblich, so ist das Pendel mittlerweile wieder in die andere Richtung ausgeschlagen, was sich am Interesse und der Präsenz teleologischer Handlungserklärungen zeigt. Vgl. Horn & Löhner 2010a.

<sup>121</sup> Unter einer ‘pro-attitude’ (Pro-Einstellung) versteht Davidson „[...] desires, wantings, urges, promptings, and a great variety of moral views, aesthetic principles, economic prejudices, social conventions, and public and private goals and values in so far as these can be interpreted as attitudes of an agent directed toward actions of a certain kind.” (Davidson 1963, 4) Pro-Einstellungen können also als Wünsche bezogen auf Typen von Handlungen verstanden werden. Der Begriff ‘desire’ oder Wunsch spielt in der gesamten handlungstheoretischen Debatte eine entscheidende Rolle. Vgl. Kap. III.5.4.

<sup>122</sup> „What is the relation between a reason and an action when the reason explains the action by giving the agent’s reasons for doing what he did? We may call such explanations *rationalizations*, and say that the reason *rationalizes* the action. In this paper I want to defend the ancient – and commonsense – position that rationalization is a species of causal explanation.” (Davidson 1963, 3)

Gründe rechtfertigen bzw. rationalisieren außerdem die Handlungen, die sie verursachen. Gründe stehen damit zugleich in einer Rationalisierungs- und einer Kausalbeziehung zu den Handlungen, die sie rechtfertigen und verursachen.<sup>123</sup> Davidsons Hauptargument für diese These ist, dass ohne die kausale Lesart der Rationalisierungsbeziehung nicht verständlich ist, was es heißt, dass eine Person *aus* einem bestimmten Grund und nicht aus einem anderen gehandelt hat. Oder anders: Echte Handlungserklärungen geben im Unterschied zu bloßen Rationalisierungen die Gründe der Person an, welche die Handlung nicht nur aus Sicht der Person rechtfertigten, sondern sie auch verursachten.<sup>124</sup> Denn es sei möglich, dass eine Person einen Grund habe, etwas zu tun, dasselbe aber aus einem anderen Grund tue.

„[A] person can have a reason for an action, and perform the action, and yet this reason not be the reason why he did it. Central to the relation between a reason and an action it explains is the idea that the agent performed the action *because* he had the reason.” (Davidson 1963, 9)

Intentionen spielen in diesem frühen Konzept von Davidson keine Rolle für die Erklärung und Explikation von Handlungen.<sup>125</sup> Über den primären Grund eines Akteurs für eine Handlung erfahre ich zugleich die Intention des Akteurs für diese Handlung.<sup>126</sup> Diese Position sollte Davidson später revidieren, auch aufgrund seiner Behandlung des Problems der Willensschwäche.<sup>127</sup>

## II.2.2 Wie ist Willensschwäche möglich?

Willensschwäche stellt innerhalb Davidsons kausaler Handlungstheorie deshalb ein grundlegendes Problem dar, weil es die Korrelation der Stärke von Gründen und Ursachen in Frage

---

<sup>123</sup> „The action on the one hand, and the belief-desire pair which give the reason on the other, must be related in two very different ways to yield an explanation. First, there must be a logical relation. [...] Second, the reasons an agent has for acting must, if they are to explain the action, be the reasons on which he acted; the reasons must have played a *causal* role in the occurrence of the action.” (Davidson 1982, 173)

<sup>124</sup> Die Unterscheidung zwischen der echten Erklärung einer Handlung und einer bloßen Rationalisierung als Bedingung der Adäquatheit für Handlungserklärungen wird auch als ‚Davidsons challenge‘ bezeichnet. Vgl. Horn & Löhner 2010a, 19.

<sup>125</sup> „If this account is correct, then acting with an intention does not require that there be any mysterious act of the will or special attitude or episode of willing. For the account needs only desires (or other pro attitudes), beliefs, and the actions themselves. There is indeed the relation between these, causal or otherwise, to be analysed, but that is not an embarrassing entity that has to be added to the world’s furniture. We would not, it is true, have shown how to *define* the concept of acting with an intention; the reduction is not definitional but ontological.” (Davidson 1978, 87f.)

<sup>126</sup> „To know a primary reason why someone acted as he did is to know an intention with which the action was done.“ (Davidson 1963, 7) „The expression ‘the intention with which James went to church’ has the outward form of a description, but in fact it is syncategorematic and cannot be taken to refer to an entity, state, dispositive, or event. Its function in context is to generate new descriptions of actions in terms of their reasons [...].” (Davidson 1963, 8)

<sup>127</sup> „At one time (about twenty-five years ago, when I wrote “Actions, Reasons, and Causes”) I thought there were no such states as intending; there were just intentional actions. This was, I now believe, an error. This is clear in the case where an intention is formed long before the intended action is performed, and even clearer in the case where the intended action is never performed. Intentions are also required to explain how complex actions are monitored and controlled.” (Davidson 1987, 106) Vgl. Moya 1990, 146.

stellt und damit die gesamte Konzeption von Rationalisierungserklärungen als Spielart von Kausalerklärungen.

„Causal theories of action are challenged by intentional actions that are contrary to the actor’s best judgment. For if reasons are causes, it is natural to suppose that the strongest reasons are the strongest causes.” (Davidson 2001, xvi)

Willensschwäche (incontinence) versteht Davidson dabei wie folgt:

„In doing x an agent acts incontinently if and only if: (a) the agent does x intentionally; (b) the agent believes there is an alternative action y open to him; (c) the agent judges that, all things considered, it would be better to do y than to do x.” (Davidson 1969a, 22)

Diese Definition von willensschwachen Handlungen weicht, wie Davidson bemerkt, in einigen Punkten von traditionellen Bestimmungen ab.<sup>128</sup> *Erstens* ist es hinreichend für Willensschwäche, dass die Person bloß gegen ein besseres *Urteil*, und nicht wider besseres *Wissen*, handelt. Aristoteles versteht dagegen willensschwache Handlungen primär als Handlungen wider besseres Wissen, und erst davon abgeleitet als Handeln wider das bessere Urteil.<sup>129</sup> *Zweitens* spielt es für diese Definition keine Rolle, in welchem Sinne die Person die alternative Handlung als besser bewertet. Dies können moralische aber auch nicht-moralische Gründe sein. Bei Aristoteles ist der Inhalt der Urteile (bzw. des Wissens) primär bezogen auf bestimmte Affekte, denen nicht im Übermaß zu folgen ist, nämlich Lust und Unlust bezüglich sinnlicher Genüsse. Nur in einer uneigentlichen Form können auch andere Affekte als Basis für Zuschreibungen von Willensschwäche dienen. Davidsons Definition ist damit bewusst neutral in Hinblick auf die moralische Dimension formuliert, was er als Vorteil gegenüber traditionellen Definitionen ansieht.<sup>130</sup> *Drittens* reicht nach Davidson ein komparatives statt eines superlativen Urteils für die Zuschreibung von Willensschwäche aus: die Person urteilt, dass eine Handlung *besser* sei als eine andere und nicht notwendigerweise, dass sie auch die *beste* sei. *Viertens* ist zu bemerken, dass auch bei Davidson die Person zumindest glauben muss, dass ihr die alternative Handlung offensteht und dass ihre willensschwache Handlung intentional ist. Diese Bedingungen sollen die Freiheit oder zumindest Ungezwungenheit der willensschwachen Handlung sicherstellen, die dann als Voraussetzung für die Verantwortlichkeitszuschreibung fungieren kann.

Davidson formuliert nun das Problem, das Willensschwäche aufwirft, als ein Trilemma: drei Aussagen, die für sich genommen plausibel sind, jedoch untereinander zu Widersprüchen führen.

„P1. If an agent wants to do x more than he wants to do y and he believes himself free to do either x or y, then he will intentionally do x if he does either x or y intentionally.

---

<sup>128</sup> Vgl. Davidson 1969a, 21f.

<sup>129</sup> Vgl. Kap. II.1.3 Eckpunkt E6.

<sup>130</sup> Vgl. Davidson 1969a, 30.

P2. If an agent judges that it would be better to do  $x$  than to do  $y$ , then he wants to do  $x$  more than he wants to do  $y$ .

P3. There are incontinent actions.” (Davidson 1969a, 23)

P1 formuliert einen Zusammenhang zwischen den Wünschen (Strebungen) einer Person und ihren intentionalen Handlungen, P2 einen zwischen den evaluativen Urteilen einer Person über Handlungsmöglichkeiten und ihren Wünschen.<sup>131</sup> Wenn eine Person nun willensschwach handelt, dann urteilt sie gemäß der Definition von Willensschwäche, dass eine Handlung  $y$  besser wäre als  $x$ , also sollte sie gemäß P2 auch einen stärkeren Wunsch haben,  $y$  als  $x$  zu tun, und gemäß P1 sollte sie demnach auch  $y$  wirklich tun, wenn sie intentional handelt. Aber laut P3 handelt die Person willensschwach und vollzieht damit die intentionale Handlung  $x$  und nicht  $y$ .

Davidsons Lösung dieses Trilemmas besteht nun darin, zwischen verschiedenen Arten, Urteile zu fällen, zu unterscheiden. Während in der Definition des willensschwachen Handelns die Person nur *prima facie* urteilt,<sup>132</sup> d.h. *hinsichtlich* einer bestimmten Menge an Gründen (all things considered), ist in P2 von unbedingten (all-out judgments) die Rede. Eine Person urteilt z.B., dass es hinsichtlich sportlicher Betätigungen erstrebenswert wäre, heute Tennis zu spielen, dass es aber hinsichtlich des schlechten Wetters nicht erstrebenswert wäre, heute Tennis zu spielen. Dieselbe Handlung kann demnach in vielerlei Hinsichten als erstrebenswert oder nicht erstrebenswert beurteilt werden. Aus diesen *prima facie* Urteilen lassen sich jedoch logisch keine unbedingten (all-out) Urteile deduktiv ableiten wie z.B. ‚Tennispielen ist heute erstrebenswert‘.<sup>133</sup> Teilweise formalisiert – pf( ) als *prima facie* Operator – könnte man nach Davidson diese praktische Überlegung folgendermaßen verstehen:

(P1) pf(Tennis spielen ist besser als zu Hause zu bleiben; sportliche Betätigung)

(P2) pf(Zu Hause zu bleiben ist besser als Tennis zu spielen; schlechtes Wetter)

(P3) pf(Tennis spielen ist besser als zu Hause zu bleiben; alles in allem)

(K) Also: Tennis spielen ist besser als zu Hause zu bleiben

In *Actions, reasons, and causes* hatte Davidson noch nicht den *prima facie* Charakter von Wünschen und anderen Pro-Einstellungen anerkannt und sie durch ein unbedingtes Urteil repräsentiert. Dadurch konnte er zur Erklärung von intentionalen Handlungen auf Intentionen verzichten, da

---

<sup>131</sup> Eine Strebung ist ein Wunsch, der sich auf eine konkrete Handlung richtet. Nach dem klassischen Wunsch-Überzeugungs-Modell verfügt die Person nicht nur über einen Wunsch nach einem Handlungsergebnis, sondern auch über eine Strebung nach einer Handlung, gegeben diesen Wunsch und Überzeugungen zu kausalen Verknüpfungen zwischen Handlungen und Handlungsergebnissen.

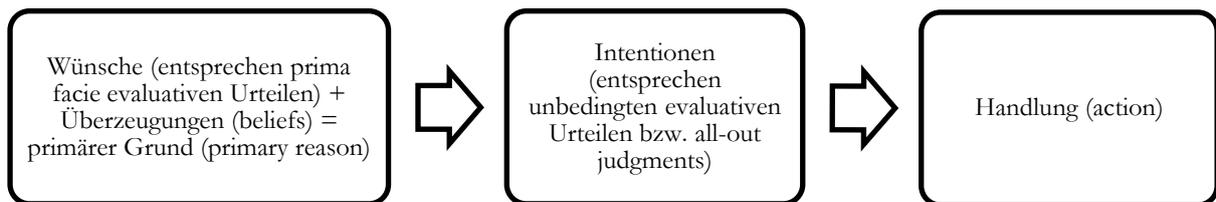
<sup>132</sup> „Let us call judgments that actions are desirable in so far as they have a certain attribute, *prima facie* judgements.” (Davidson 1978, 98)

<sup>133</sup> Davidson modelliert diese Art praktischer Schlüsse nach dem Vorbild von Schlüssen aus probabilistischen Evidenzen.

aus diesen unbedingten Urteilen ein weiteres unbedingtes Urteil abgeleitet werden konnte, das die jeweilige Handlung repräsentierte bzw. ausdrückte.

„In ‘How is weakness of the will possible?’ I gave up the deductive model, and changed what I took to be the logical character of the propositional content of a want or desire. The premiss someone has who want to see what is on the next page is not ‘Any act of mine that has a good chance of letting me see what is on the next is desirable’ but something more like ‘Any act of mine is desirable in so far as it has a good chance of letting me see what is on the next page’. This premiss does indeed combine with the contents of a belief to the effect that turning over a new leaf has the desirable characteristic, but what can be deduced is only ‘Turning over a new leaf is desirable in so far as it has a good chance of letting me see what is on the next page’. Such a conclusion, I argued, could not correspond to or be identified with an action since it is compatible with the agent knowing that the action (because of other characteristics) is highly undesirable. Actions, I urged, correspond to ‘all-out’ judgements like ‘Turning over the page is desirable’, and such judgements do not follow by ordinary logic from the premises provided by our desires and beliefs.” (Davidson 1985a, 196)

Auf diesem Weg gelangte Davidson dazu, Intentionen in seine kausale Handlungstheorie als eigenständige Spielart unbedingter Urteile zu integrieren, auf die Handlungen kausal folgen oder diese begleiten. Diese Idee sollte er in seinem Aufsatz *Intending* von 1978 weiter ausbauen und verfolgen. Strukturell kann diese Veränderung so ausgedrückt werden:



Während die willensschwache Person also nur prima facie urteilt, dass die Handlung y besser zu tun wäre als x, urteilt sie unbedingt, dass x besser zu tun ist als y, d.h. sie formt die Intention x zu tun. Das prima facie Urteil der willensschwachen Person ist jedoch nicht ein einfaches prima facie Urteil, sondern eines, das in Anbetracht aller zur Verfügung stehenden Gründe aus der Sicht des Akteurs gefällt wird, ein so genanntes Alles-in-allem-Urteil (all things considered judgment). Dies ist wichtig, da dieses besondere Urteil nach Davidson mit speziellen Rationalitätsanforderungen verknüpft ist. Das Alles-in-allem-Urteil ist das beste Urteil, das dem Akteur zu einer gewissen Zeit hinsichtlich der ihm verfügbaren Gründe zur Verfügung steht. Die willensschwache Person verstößt gegen ein Prinzip der Anwendung praktischer Rationalität, das Davidson Kontinenzprinzip (principle of continence) nennt:

„But what, on this analysis, is the fault in incontinence? The akrates does not, as is now clear, hold logically contradictory beliefs, nor is his failure necessarily a moral failure. What is wrong is that the incontinent man acts, and judges, irrationally, for this is surely what we must say of a man who goes against his own best judgement. Carnap and Hempel have argued that there is a principle which is no part of the logic of inductive (or statistical) reasoning, but is a directive the rational man will accept. It is the requirement of total evidence for inductive reasoning: give your credence to the hypothesis supported by all available relevant evidence. There is, I suggest, an analogous principle the rational man will accept in applying practical reasoning: perform the action judged best on the basis of all available relevant reasons. It would be appropriate to call this the principle of continence.” (Davidson 1969a, 41)

Obwohl die willensschwache Person einen Grund hat für ihre willensschwache Handlung – es liegen Wünsche und andere mentale Zustände vor, die sich als prima facie Urteile für die Ausführung der Handlung repräsentieren lassen – hat sie keinen Grund, gegen das Kontinenzprinzip zu verstoßen. Es bleibt daher immer ein Rest an Unverständlichkeit bestehen in dem Versuch des Akteurs und seiner Umwelt, sein Verhalten zu verstehen, d.h. es zu rationalisieren.<sup>134</sup>

Die logische Möglichkeit von Willensschwäche besteht nach Davidson darin, dass prima facie Urteile und unbedingte Urteile einander nicht widersprechen. Es bleibt die Frage, wie es faktisch möglich ist, dass eine Person zwar prima facie urteilt, dass es alles in allem besser wäre, die Handlung y zu tun, und dennoch unbedingt urteilt, dass x zu tun besser sei. Davidson führt dazu das Konzept von Gründen ein, die Ursachen sind, ohne ihre Wirkungen zu rationalisieren.

„There is no paradox in supposing a person sometimes holds that all that he believes and values supports a certain course of action, when at the same time those same beliefs and values cause him to reject that course of action. If r is someone's reason for holding that p, then his holding that r must be, I think, a cause of his holding that p. But, and this is what is crucial here, his holding that r may cause his holding that p without r being his reason; indeed, the agent may even think that r is a reason to reject p.” (Davidson 1969a, 41)

Doch das Konzept von Gründen, die Ursache für andere mentale Zustände oder Handlungen sind, ohne diese zu rationalisieren, benötigt eine Theorie, die zeigt, wie eine Person über solche Gründe verfügen kann. Dies soll Davidsons Teilungstheorie des Geistes leisten.

### II.2.3 Paradoxien der Irrationalität

In seinem Aufsatz *Paradoxes of Irrationality* von 1982 erweitert Davidson seine Beschäftigung mit irrationalen Phänomenen und vertieft sein Konzept von Gründen, die Ursachen sind für andere mentale Zustände, ohne diese zu rationalisieren. Dabei stößt er auf die Paradoxie der Irrationalität: je mehr ein irrationales Phänomen verstanden wird, indem es in ein intelligibles Netz von Gründen eingegliedert wird, desto mehr verliert dieses seinen irrationalen Charakter.<sup>135</sup>

„The underlying paradox of irrationality, from which no theory can entirely escape, is this: if we explain it too well, we turn it into a concealed form of rationality; while if we assign incoherence too glibly, we merely compromise our ability to diagnose irrationality by withdrawing the background of rationality needed to justify any diagnosis at all.” (Davidson 1982, 184)

Angewendet auf das Phänomen der Willensschwäche lässt sich das Dilemma folgendermaßen verstehen: Entweder die willensschwache Handlung hat Gründe, die sie bloß verursachen, ohne sie zu rationalisieren. Dann ist fraglich, wieso diese Handlung intentional ist, sprich überhaupt

---

<sup>134</sup> „What is special in incontinence is that the actor cannot understand himself: he recognizes, in his own intentional behaviour, something essentially surd.” (Davidson 1969a, 42)

<sup>135</sup> Vgl. in diesen Zusammenhang die allgemeine Diskussion um den Interpretationismus in der Philosophie des Geistes bei Heil 2004, ch. 10.

eine Handlung und nicht ein bloßes Ereignis. Oder die Gründe für die willensschwache Handlung rationalisieren die Handlung, dann ist unklar, weshalb diese irrational bzw. problematisch ist.<sup>136</sup> Davidsons Lösung dieses Dilemmas besteht darin, eine Teilungsthese des Geistes nach Vorbild der Freudschen psychoanalytischen Tradition einzuführen. Innerhalb eines mentalen Teiles rationalisiert der Grund die willensschwache Handlung und verursacht, ohne zu rationalisieren, in einem anderen mentalen Teil eine Blockade der Anwendung des Kontinenzprinzips. Die Person handelt daher intentional und rational hinsichtlich des einen mentalen Teils, aber irrational hinsichtlich des anderen, der dem Kontinenzprinzip folgt und damit auch dem aus eigener Perspektive gültigen Maßstab für praktische Rationalität genügt. Mentale Teile sind für Davidson semi-autonome Strukturen, die wiederum intern strukturiert sind in miteinander zusammenhängende Überzeugungen, Wünsche und andere mentale Zustände. Die Grenzen eines mentalen Teils sind die fehlenden rationalen Beziehungen zwischen verschiedenen mentalen Zuständen, d.h. dort, wo ein mentaler Zustand lediglich Ursache, aber nicht Grund für einen anderen ist.

## **II.2.4 Probleme & Kritik I: Davidsons kausale Handlungstheorie**

Die Kritik an Donald Davidsons kausaler Handlungstheorie ist vielfältig und mittlerweile äußerst umfangreich. Viele der besprochenen Einwände sind dabei gegen kausale Handlungstheorien allgemein gerichtet und betreffen demnach nicht bloß seine eigene Konzeption.<sup>137</sup> Ich beschränke mich hier auf zentrale Einwände, darunter das Logische-Beziehungs-Argument, das Problem abweichender Kausalketten, das Gesetze-Argument, das Dispositions-Argument und das Handlungen-ohne-Akteure-Argument.

### **a) Das Logische-Beziehungs-Argument**

Das Argument der logischen Verknüpfung<sup>138</sup> (logical connection argument) besagt, dass zwischen Handlungen und ihren mentalen Ursachen logische Verbindungen bestehen, was eine kausale Analyse unmöglich mache, da eine kausale Verbindung nur zwischen zwei logisch voneinander

---

<sup>136</sup> „So we face the following dilemma: if we think of the cause in a neutral mode, disregarding its mental status as a belief or other attitude – if we think of it merely as a force that works on the mind without being identified as part of it – then we fail to explain, or even describe, irrationality. Blind forces are in the category of the non-rational, not the irrational. So, we introduce a mental description of the cause, which thus makes it a candidate for being a reason. But we still remain outside the only clear pattern of explanation that applies to the mental, for that pattern demands that the cause be more than a candidate for being a reason; it must *be* a reason, which in the present case it cannot be. For an explanation of a mental effect we need a mental cause that is also a reason for this effect, but, if we have it, the effect cannot be a case of irrationality. Or so it seems.” (Davidson 1982, 180)

<sup>137</sup> Zu diesen Einwänden und deren Bezeichnungen vgl. Beckermann 1977.

<sup>138</sup> Zu diesem Argument vgl. Beckermann 1977; Beckermann 1985; Davis 1979, 87ff. Vertreten wurde es unter anderem von Abraham Melden und George H. von Wright.

unabhängigen Ereignissen (oder allgemein Entitäten) bestehen könne.<sup>139</sup> Davidson erwidert, dass logische Verknüpfungen nur zwischen *Beschreibungen* von Entitäten, nicht zwischen den Entitäten selbst bestehen, bzw. bei Handlungen zwischen Beschreibungen von Ereignissen, und nicht den Ereignissen selbst.

„In any case there is something very odd in the idea that causal relations are empirical rather than logical. What can this mean? Surely not that every true causal statement is empirical. For suppose '*A* caused *B*' is true. Then the cause of *B* = *A*; so substituting, we have 'The cause of *B* caused *B*', which is analytic. The truth of a causal statement depends on what events are described; its status as analytic or synthetic depends on how the events are described.“ (Davidson 1963, 14)

Obwohl Davidson die Rationalisierung als Kausalerklärung versteht, trennt er dennoch strikt zwischen der kausalen und der normativen Ebene. Die kausale Relation besteht zwischen Ereignissen unabhängig ihrer Beschreibungen, die normative Beziehung – ‚vernünftig im Lichte von Gründen‘ – zwischen bestimmten Beschreibungen der Ereignisse, d.h. zwischen den durch sie ausgedrückten Propositionen. Das Problem der abweichenden Kausalketten ist das Resultat dieser strikten Trennung: durch die strukturelle Trennung der beiden Ebenen kann die kausale Beziehung zwischen Gründen und Handlungen intakt bleiben, während die normative bzw. rechtfertigende Beziehung zwischen Gründen und Handlungen verlorenght.

„There is a *split* between causality and rationality: reasons do not cause actions in virtue of their justifying or rationalizing them, but in virtue of their instantiating a strict law; and reasons rationalize actions in a way sharply separated from their causal aspect, namely by virtue of having descriptions which cohere appropriately with the description or descriptions under which the agent saw his actions.“ (Moya 1990, 112)

Der Preis dafür, dass das Logische-Beziehungs-Argument nicht greift, ist demnach sehr hoch und über die strikte Trennung der normativen und der kausalen Ebene erkaufte.

### **b) Abweichende Kausalketten**

Das Problem abweichender Kausalketten (wayward bzw. deviant causal chains) ist ein Grundproblem für kausale Handlungstheorien.<sup>140</sup> Diese verstehen Handlungen als distinkt von anderen Ereignissen bzw. bloßen Vorkommnissen durch die Art ihrer Ursachen. Handlungen

---

<sup>139</sup> Vgl. Davidson 1963, 13–15. Geert Keil beschreibt das Argument folgendermaßen: „(P<sub>1</sub>) Zwischen Handlungsgründen (oder: Motiven, Intentionen, Wünschen) und der durch sie rationalisierten Handlung besteht eine »logische« Beziehung. (P<sub>2</sub>) Das Bestehen einer logischen Beziehung und das einer kausalen Beziehung zwischen denselben Relata schließen einander aus, *ergo*: (K) Zwischen den in (P<sub>1</sub>) genannten Entitäten kann keine kausale Beziehung bestehen.“ (Keil 2000, 59) Dasselbe bei Beckermann: „Die geistigen Zustände, auf die in mentalen Erklärungen Bezug genommen wird, können nicht die Ursachen der mit ihrer Hilfe erklärten Handlungen sein, da die Beziehung zwischen Ursache und Wirkung eine empirische ist, zwischen den geistigen Zuständen eines Handelnden und seinen Handlungen aber keine empirische, sondern eine logische Beziehung besteht.“ (Beckermann 1977, 64)

<sup>140</sup> Für die mittlerweile sehr umfangreiche Diskussion zu diesem Thema vgl. Brand 1984, Kap. 1; Mele 1987b; Mele 1992, 208–210; Frankfurt 1978; Moya 1990, 115–128; Keil 2000, Kap. I.4; Searle 1983, ch. 3–4; Thalberg 1984; Mele & Moser 1994, 44–52; Enç 2004; Mitchell 1982; Goldman 1970, Kap. 3; Stout 2010; Pears 1975.

(oder intentionale Handlungen)<sup>141</sup> werden verursacht durch Wünsche und entsprechende Überzeugungen oder darauf basierende Intentionen. Es ist jedoch möglich, dass ein Ereignis durch Gründe verursacht wird, das entsprechend intendierte Resultat aber auf eine solche Weise zustande kommt, dass niemand das Ereignis als (intentionale) Handlung beschreiben würde.<sup>142</sup>

„Instances of deviant or “wayward” causal chains raise difficulties for causal analyses of action and intentional action. [...] The alleged problem about *intentional* action in particular is that whatever psychological causes are deemed both necessary and sufficient for a resultant action’s being intentional, cases can be described in which, due to a deviant causal connection between the favored psychological antecedents and a pertinent action to which they lead, that action is not intentional.” (Mele 1992, 208)

Ein Beispiel: Daniel sitzt in einer Bar und möchte noch eine Flasche Bier bekommen. Er versucht, der Kellnerin ein entsprechendes Zeichen zu geben, indem er mit der leeren Flasche in der Hand winkt. Die Kellnerin sieht ihn nicht, im Gegensatz zum Hund der Kellnerin, der, was Daniel nicht wusste, auf dieses Zeichen trainiert ist und Daniel eine weitere Flasche Bier bringt. Der Wunsch nach einer weiteren Flasche und die Überzeugung, mit der leeren Flasche zu winken würde ihm ein neues Bier beschaffen, verursachten, dass Daniel mit der leeren Flasche in der Hand winkte, was wiederum verursachte, dass er eine neue Flasche bekam. Er bekam aber die neue Flasche auf eine Weise, die Daniel nicht vorhergesehen hatte und die er bestimmt nicht intendiert hatte, da er gar nicht wusste, dass der Hund so etwas tun könnte. Die kausale Verknüpfung könnte natürlich noch viel absurdere Wendungen nehmen, solange das intendierte Resultat eintritt, bleibt das Problem bestehen, dass es nicht auf die richtige Art und Weise verursacht wurde, obwohl die Gründe, welche die Handlung rationalisieren, sie auch verursacht haben. Strukturell lässt sich das Argument so verstehen:

- (P1) Eine intentionale Handlung ist die Wirkung eines sie verursachenden Grundes, der gleichzeitig die Handlung rationalisiert (kausale Handlungstheorie)
- (P2) Gründe können Handlungen verursachen und rationalisieren, welche nicht intentional sind (abweichende Kausalketten)
- (K) Also: Nicht jede Wirkung, die durch einen Grund verursacht wird, der sie auch rationalisiert, ist eine intentionale Handlung.

---

<sup>141</sup> Die Formulierung hängt davon ab, ob Intentionalität als Bestandteil des Begriffs einer Handlung verstanden wird oder nicht. Nach Davidson gibt es keine eigene Klasse von ‚intentionalen Handlungen‘, d.h. es gibt keine neutrale Klasse von ‚Handlungen‘, die in intentionale und nicht-intentionale Handlungen zerfällt, sondern Handlungen sind immer intentional unter mindestens einer Beschreibung, unter anderen nicht. „It is a mistake to suppose there is a class of intentional actions: if we took this tack, we should be compelled to say that one and the same action was both intentional and not intentional.” (Davidson 1971, 46)

<sup>142</sup> Ein klassisches Beispiel für diese Möglichkeit lieferte Roderick Chisholm 1966. Vgl. auch das Beispiel von Daniel Bennett, das Davidson anführt: „A man may try to kill someone by shooting at him. Suppose the killer misses his victim by a mile, but the shot stampedes a herd of wild pigs that trample the intended victim to death. Do we want to say the man killed his victim *intentionally*? The point is that not just any causal connection between rationalizing attitudes and a wanted effect suffices to guarantee that producing the wanted effect was intentional. The causal chain must follow the right sort of route.” (Davidson 1973, 78)

Die Kausalreihe kann nicht nur in den Konsequenzen von Handlungen abweichen, sondern bereits zwischen Wünschen und Überzeugungen und dem Handlungsbeginn.<sup>143</sup> So könnte der Wunsch und die entsprechende Überzeugung, ein Bier bei der Kellnerin zu bestellen, Daniel so nervös machen, dass er panisch mit seiner Bierflasche in der Luft winkt, was wiederum den Hund dazu veranlasst, ein weiteres Bier zu holen. Der Phantasie ist freien Lauf gelassen, die kausale Verknüpfung so absurd wie möglich zu gestalten, solange die intendierte Handlung verursacht wird und das intendierte Handlungsergebnis eintritt. Davidson hat dieses Problem erkannt:

„There must be such rationalizing beliefs and desires if an action is done for a reason, but of course the presence of such beliefs and desires when the action is done does not suffice to ensure that what is done is done with the appropriate intention, or even with any intention at all. Someone might want tasty stew and believe sage would do the trick and put in sage thinking it was parsley; or put in sage because his hand was joggled. So we must add that the agent put in the sage because of his reasons. This 'because' is a source of trouble; it implies, so I believe, and have argued at length, the notion of cause. But not any causal relation will do, since an agent might have attitudes and beliefs that would rationalize an action, and they might cause him to perform it, and yet because of some anomaly in the causal chain, the action would not be intentional in the expected sense, or perhaps in any sense. We end up, then, with this incomplete and unsatisfactory account of acting with an intention: an action is performed with a certain intention if it is caused in the right way by attitudes and beliefs that rationalize it.” (Davidson 1978, 87)

Davidson ist skeptisch gegenüber der Aussicht, eine reduktive kausale Analyse des Begriffs des intentionalen Handelns ohne eine entsprechende *ceteris paribus* Klausel wie ‚in der richtigen Weise‘ (*in the right way*) einzufügen.<sup>144</sup> Der Grund besteht nach Davidson darin, dass wenn wir dies tun könnten, d.h. alle Bedingungen explizit angeben könnten, aus denen kausal hinreichend intentionales Handeln folgt, wir ein striktes psycho-physisches Gesetz formuliert hätten, was wiederum seinen Grundsätzen des mentalen Holismus und anomalen Monismus in der Philosophie des Geistes widersprechen würde.<sup>145</sup> Es wurden einige Versuche unternommen, den Einwand basierend auf abweichende Kausalkette abzuschwächen bzw. zu verhindern.<sup>146</sup> Alvin Goldman vertritt dabei die Ansicht, dass die Philosophie gar nicht die Aufgabe habe, die kausale Analyse so zu verfeinern, dass sie nicht bloß notwendige, sondern auch hinreichende Bedingungen für das Vorliegen intentionaler Handlungen angibt:

---

<sup>143</sup> Vgl. die Unterscheidung zwischen ‚antecedential‘ und ‚consequential waywardness‘ bei Brand 1984, 18ff. und zwischen ‚primary‘, ‚secondary‘ und sogar ‚tertiary waywardness‘ bei Mele 1987b.

<sup>144</sup> „The point is that not just any causal connection between rationalizing attitudes and a wanted effect suffices to guarantee that producing the wanted effect was intentional. The causal chain must follow the right sort of route.” (Davidson 1973, 78) „Although intentional actions are caused by intentions, it is not enough to insure that an action was performed with a certain intention that it was caused by that intention. [...] Deviant causal chains [...] present a problem in the explanation of action, since we would like to be able to say what the conditions are that must be satisfied if an action is to be intentional. Several clever philosophers have tried to show how to eliminate the deviant causal chains, but I remain convinced that the concepts of event, cause, and intention are inadequate to account for intentional action.” (Davidson 1987, 106)

<sup>145</sup> Vgl. Keil 2000, 98ff. Bishop ist dagegen der Meinung, dass eine kausale Handlungstheorie gar nicht den Anspruch erheben muss, eine reduktive Explikation des Handlungsbegriffes zu geben, sondern lediglich notwendige und hinreichende Bedingungen für das faktische Stattfinden von Handlungen. Vgl. Bishop 1989, 176ff.

<sup>146</sup> Ich werde hier nicht weiter auf die verschiedenen Strategien eingehen. Vgl. Keil 2000, Kap. I.4.

„The difficult question, of course, is: precisely what is this “characteristic” mode of causation by which wants and beliefs cause intentional action? To this question, I confess, I do not have a fully detailed answer. But neither do I think that it is incumbent on me, *qua* philosopher, to give an answer to this question. A complete explanation of how wants and beliefs lead to intentional acts would require extensive neurophysiological information, and I do not think it is fair to demand of a *philosophical* analysis that it provide this information.” (Goldman 1970, 62)

„To help clarify my position, compare the analysis of intentional action with the analysis of seeing. An analysis of what it is for someone to see a physical object must, I believe, require there to be a causal connection between the object and the sensory contents of the percipient. But not *any* such causal connection or other will do, for there are some causal processes leading from an object to visual sensory impressions which would *not* count as the person’s *seeing* the object. What is needed, then, is a specification of the *appropriate* causal process which constitutes seeing. But is it legitimate to expect a philosophical analysis of seeing to specify this process in detail? I think not.” (Goldman 1970, 62f.)

Der Vergleich mit der kausalen Analyse der Wahrnehmung ist natürlich kein Argument, sondern zeigt lediglich, dass abweichende Kausalketten für jede Art der kausalen Analyse intentionaler Beziehungen zu Schwierigkeiten führen. Goldman im Gegensatz zu Davidson unterschätzt die Sprengkraft des Einwandes basierend auf abweichender Kausalketten für kausale Handlungstheorien: wenn der Anspruch aufgegeben wird, hinreichende Bedingungen für das Vorliegen von intentionalen Handlungen anzugeben, verliert die kausale Analyse ihren ‚Witz‘. Schließlich kann jede andere Handlungstheorie kausale Elemente im Sinne notwendiger Bedingungen für die Analyse von Handlungen bereitwillig aufnehmen ohne damit den Begriff der Handlung kausal zu explizieren. Die Konsequenz dieser Abschwächungsversuche ist demnach, die Reinheit der kausalen Handlungstheorie zu verwässern und normative Elemente in die kausale Analyse einzufügen. Dies sollte nicht verwundern, da ‚auf die richtige Weise‘ eindeutig normativ zu verstehen ist und nicht rein deskriptiv.<sup>147</sup>

### **c) Das Gesetze-Argument**

Das Gesetze-Argument besagt, dass Rationalisierungen, d.h. die Angabe von Gründen von Akteuren für ihre Handlungen, keine kausalen Erklärungen von Handlungen sein können, da für kausale Erklärungen strikte Gesetze angeführt werden müssen, während Rationalisierungen ohne die Angabe solcher strikten Gesetze funktionieren.<sup>148</sup> Davidson bestreitet die zugrundeliegende These damit, dass wir wahre singuläre Kausalurteile auch ohne Kenntnis der sie subsumierenden Gesetze fällen können.

---

<sup>147</sup> „There was, after all, a point in the clause ‘in the right way’. But the place of this clause was wrong. It is not the causal chain that has to operate in the right way: it is the way the agent follows his plan or intention that has to be right. [...] Following a plan or fulfilling an intention are subject to normative constraints, and the requirement of correctness is perfectly adequate here, whereas a causal chain is not subject to those constraints, and it is odd to require it to be right. The initial appeal of this clause derives in fact from the normative structure of intentional action.” (Moya 1990, 126)

<sup>148</sup> Vgl. Davidson 1963, 15–17. „In brief, laws are involved essentially in ordinary causal explanations, but not in rationalizations.” (Davidson 1963, 15)

„The most primitive explanation of an event gives its cause; more elaborate explanations may tell more of the story, or defend the singular causal claim by producing a relevant law or by giving reasons for believing such exists. But it is an error to think no explanation has been given until a law has been produced. Linked with these errors is the idea that singular causal statements necessarily indicate, by the concepts they employ, the concepts that will occur in the entailed law. Suppose a hurricane, which is reported on page 5 of Tuesday's Times, causes a catastrophe, which is reported on page 13 of Wednesday's Tribune. Then the event reported on page 5 of Tuesday's Times caused the event reported on page 13 of Wednesday's Tribune. Should we look for a law relating events of these kinds! It is only slightly less ridiculous to look for a law relating hurricanes and catastrophes.” (Davidson 1963, 17)

Nach Davidson bestehen kausale Relationen zwischen Ereignissen unabhängig von der Art und Weise, wie wir diese Ereignisse beschreiben. Die Wahrheit eines Kausalurteils hängt demnach davon ab, auf welche Ereignisse referiert wird – singuläre Kausalurteile sind extensional –, während Gesetzaussagen Ereignisse unter bestimmten Beschreibungen miteinander verknüpfen, also intensional sind, wodurch deren Wahrheit von der Art und Weise abhängen, wie die involvierten Ereignisse beschrieben werden. Nach Davidsons anomalen Monismus sind psychophysische Gesetzaussagen sogar unmöglich: Obwohl wir wissen, dass es Gesetze gibt, welche die mentalen Ursachen mit Handlungen verknüpfen, können wir sie nicht formulieren.<sup>149</sup> Obwohl also nach Davidson Rationalisierungen kein explizites Gesetz anführen, welche das singuläre Kausalurteil – ‚Die primären Gründe X verursachten Handlung Y‘ – subsumiert, sind sie dennoch Kausalerklärungen, da die in der Rationalisierung referierten mentalen und körperlichen Ereignisse in einer Kausalbeziehung stehen. Trotz fehlender strikter Gesetzaussagen vertritt Davidson also eine ereigniskausalistische Auffassung – Kausalbeziehungen bestehen ausschließlich zwischen Ereignissen – gepaart mit einem nomologischen Verständnis von Kausalität – wahre Kausalurteile lassen sich unter ein Gesetz subsumieren, das Ursache und Wirkung in einer strikten, empirisch gehaltvollen Aussage miteinander verknüpft.<sup>150</sup> Es lassen sich nämlich wahre singuläre Kausalurteile fällen, ohne das entsprechende Gesetz anzugeben, da es ausreicht, dass wir gute Gründe dafür haben anzunehmen, dass es ein solches Gesetz gibt, welches das singuläre Kausalurteil subsumiert.

„Causality and identity are relations between individual events no matter how described. But laws are linguistic; and so events can instantiate laws, and hence be explained or predicted in the light of laws, only as those events are described in one or another way. [...] The principle of the nomological character of causality must be read carefully: it says that when events are related as cause and effect, they have descriptions that instantiate a law. It does not say that every true singular statement of causality instantiates a law.” (Davidson 1970, 215)

Davidson behauptet also zugleich, dass Rationalisierungen kausal interpretiert werden sollten, sie aber keine Gesetze explizit anführen müssen, um Handlungen erklären zu können. Es reichen gute Gründe für die Vermutung, dass solche Gesetze existieren, obwohl wir durch den

---

<sup>149</sup> Vgl. Davidson 1970.

<sup>150</sup> Das Prinzip des nomologischen Charakters von Kausalität nach Davidson: „Events related as cause and effect fall under strict deterministic laws.” (Davidson 1970, 208) Vgl. Keil 2000, Kap. I.5.

holistischen Charakter mentaler Prädikate zugleich wissen können, dass wir solche Gesetze niemals explizit angeben können. Während Davidson das Gesetze-Argument unterläuft, indem er das Fällen singulärer kausaler Urteile unabhängig von expliziten Gesetzesaussagen macht, entschärfen nomologische Varianten kausaler Handlungstheorien das Argument durch die explizite Formulierung der entsprechenden Gesetzesaussagen. So hat Paul M. Churchland in seinem Aufsatz *The logical character of action-explanations* (1970) ein solches striktes Handlungsgesetz formuliert:

„ $L_1$ : (X) ( $\phi$ ) ( $A$ ) (If [1] X wants  $\phi$ , and [2] X believes that  $A$ -ing is a way for him to bring about  $\phi$  under those circumstances, and [3] there is no action believed by X to be a way for him to bring about  $\phi$ , under the circumstances, which X judges to be as preferable to him as, or more preferable to him than,  $A$ -ing, and [4] X has no other want (or set of them) which, under the circumstances, overrides his want  $\phi$ , and [5] X knows how to  $A$ , and [6] X is able to  $A$ , then [7] X  $A$ -s.)” (Churchland 1970, 221f.)

Kern dieses Handlungsgesetzes bildet die Angabe eines Wunsch-Überzeugungs-Paares in den Bedingungen 1 und 2, welche die Handlung rationalisiert. Die restlichen Bedingungen dienen dazu, mögliche Einwände und Abweichungen von diesem Grundschema explizit in das Handlungsgesetz aufzunehmen und so auszuschließen. So handeln Menschen nicht immer sofort, wenn sie X möchten und gleichzeitig wissen, welche Handlung X hervorbringen würde, denn es könnte ja sein, dass die Person etwas anderes mehr erstrebt. Mit Hinblick auf Davidsons Konzeption könnte man Churchlands Unterfangen auch folgendermaßen verstehen: Die Angabe des Handlungsgesetzes soll zeigen, wie aus vorliegenden Wünschen und Überzeugungen, repräsentiert durch prima facie Urteile, durch Hinzufügung weiterer Bedingungen das Eintreten von Handlungen deduktiv abgeleitet werden kann. Während Davidson bestreitet, dass ein solches Gesetz formulierbar ist, der Schluss von prima facie evaluativen Urteilen auf unbedingte Urteile – von Wünschen und Überzeugungen zu Intentionen – nicht deduktiv, sondern induktiv verstanden werden müsste, formuliert Churchland genau ein solches Gesetz. Fraglich ist natürlich, ob dieses Gesetz tatsächlich den Anforderungen an wissenschaftliche Gesetzesaussagen genügt.<sup>151</sup> Für das Handlungsgesetz  $L_1$  spielt Willensschwäche die Rolle eines Gegenbeispiels: Eine Person handelt willensschwach, wenn alle Bedingungen 1–6 erfüllt sind und dennoch 7 nicht eintritt, d.h. die Person nicht die Handlung vollzieht, welche das Gesetz erwarten lässt. Dies führt natürlich zu der bereits erwähnten argumentativen Pattstellung: Während für die einen das Phänomen der Willensschwäche ein Grund darstellt, an dem Handlungsgesetz zu zweifeln, ist für die anderen die Plausibilität des Handlungsgesetzes ein Grund, an der Existenz von Willensschwäche oder bestimmten begrifflichen Fassungen derselben zu zweifeln.<sup>152</sup> Die Plausibilität des Gesetze-Argumentes hängt natürlich von der

---

<sup>151</sup> Für Kritik vgl. Keil 2000, Kap. I.3.

<sup>152</sup> Zu Churchlands Umgang mit dem Problem der Willensschwäche vgl. Churchland 1970, 223ff.

Plausibilität des zugrundeliegenden Modells wissenschaftlicher Erklärungen nach dem deduktiv-nomologischen Verfahren ab. An diesem wurde Kritik geübt, die hier nicht weiter verfolgt werden kann. Davidson selbst hat das Verhältnis seiner nicht-nomologischen Konzeption der kausalen Handlungstheorie zu nomologischen Varianten wie die Churchlands<sup>153</sup> und Hempels<sup>154</sup> entspannt gesehen:

„The difference in our accounts, if there really is one, concerns the exact way in which *laws* are involved when we explain actions by mentioning the agent's reasons. There is a weak sense in which laws may said to be involved which is not in dispute. Hempel holds, and I agree, that if *A* causes *B*, there must be descriptions of *A* and *B* which show that *A* and *B* fall under a law. This is a weak thesis, because if this is the only sense in which laws must be invoked in reason explanations, someone might explain an action by giving the agent's reasons while having no idea what the relevant law was. I have argued that a causal relation implies the existence of strict laws belonging to a closed system of laws and ways of describing events, and that there are no such laws governing the occurrence of events described in psychological terms; we seldom if ever know how to describe actions or their psychological causes in such a way as to allow them to fall under strict laws. It would follow that we can explain actions by reference to reasons without knowing laws that link them. Hempel, if I am right, believes all explanation requires reference, oblique or direct, to relevant known empirical laws. In that case, in order to explain events we must describe them in a way that reveals how laws are applicable. This sounds like a forthright conflict, but in fact I'm not sure it is.” (Davidson 1976, 262)

Die Diskussion um den Status von Rationalisierungen – kausal oder nicht-kausal; wissenschaftlich oder nicht-wissenschaftlich; reduzierbar oder nicht-reduzierbar usw. – als Erklärungen von Handlungen kann hier nicht weiter verfolgt werden.<sup>155</sup> Das Argument stellt von sich aus jedoch keinen entscheidenden Einwand gegen kausale Handlungstheorien dar, da unter anderem Davidsons Konzept den Anspruch erhebt, die Eigenständigkeit von Rationalisierungen gegenüber anderen wissenschaftlichen Erklärungen zu bewahren und Rationalisierungen zugleich kausal zu interpretieren. Das Argument ‚nicht kausal weil nicht nomologisch‘, auf dem der Gesetzes-Einwand beruht, wird dadurch prinzipiell unterlaufen.

#### **d) Das Dispositions-Argument**

Dieses Argument besagt, dass primäre Gründe keine Ursachen für Handlungen sein können, da nur Ereignisse Ursachen anderer Ereignisse sein können, primäre Gründe im Sinne Davidsons aber Dispositionen und keine Ereignisse seien.<sup>156</sup> Davidson stimmt zu, dass Dispositionen alleine keine Ursachen sein können, aber dass der Erwerb oder das Auslösen einer Disposition ein Ereignis sein kann. Außerdem müssen wir in den meisten Fällen nur besser hinsehen, um ein entsprechendes Ereignis zu finden, das Ursache für eine gewisse Handlung war.

---

<sup>153</sup> Zu Churchland vgl. Davidson 1973, 77.

<sup>154</sup> Davidson bezieht sich hier auf Hempel 1961–62.

<sup>155</sup> Für die Diskussion vgl. Baker 1995, 126ff.; Horn & Löhrer 2010b.

<sup>156</sup> „Primary reasons consist of attitudes and beliefs, which are states or dispositions, not events; therefore they cannot be causes.” (Davidson 1963, 12)

„In many cases it is not difficult at all to find events very closely associated with the primary reason. States and dispositions are not events, but the onslaught of a state or disposition is.“ (Davidson 1963, 12)

Ergänzt wird das Dispositions-Argument oft mit einem Hinweis auf spontane und routinierte Handlungen sowie Unterlassungen.<sup>157</sup> Bei spontanen Handlungen bestehe nicht genügend Zeit, um die erforderlichen mentalen Ereignisse hervorzurufen, auch seien diese den Akteuren nicht bewusst, und daher zweifelhaft, ob sie überhaupt vorliegen. Ebenso brauche es bei routinierten Handlungen keine vorhergehenden mentalen Ereignisse, um sie auszuführen. Schließlich seien bei Unterlassungen fraglich, worin das entsprechende Ereignis bestehe, das als Handlung bzw. Unterlassung beschrieben werde. Nach der kausalen Handlungstheorie Davidsons muss es ein solches Ereignis geben, da Handlungen allgemein Ereignisse *sind*.

Der Einwand spontaner und routinierter Handlungen kann Davidson mit dem Hinweis erwidern, dass die mentalen Ereignisse, die seine Theorie in diesen Fällen voraussetzen, dem Akteur nicht explizit bewusst sein müssen, um dennoch wirksam zu sein. Davidson ist sicherlich nicht der Meinung, dass jeder intentionalen Handlung eine explizite *Überlegung* der Person vorausgehen müsste, die mit einer expliziten *Entscheidung* für eine Handlungsoption endete. Ein Beispiel: Einer Person wird ein Ball zugeworfen, den sie fängt. Die Person hat sicherlich nicht überlegt, ob sie den Ball fangen soll oder nicht, oder sich explizit dazu entschieden, den Ball auch zu fangen. Innerhalb von Davidsons Theorie können wir diesen Fall einer spontanen Handlung so verstehen: ein dispositionaler Wunsch nach körperlicher Unversehrtheit in Gemeinschaft mit dispositionalen Überzeugungen über die Situation verursachen gemeinsam mit einer spontan auftretenden Überzeugung, gebildet auf der Basis von Wahrnehmungen des sich nähernden Balles, die intentionale Handlung, den Ball zu fangen. Das mentale Ereignis, welche die bestehenden Dispositionen auslöst, ist in diesem Fall die durch die Wahrnehmung verursachten spontanen Überzeugungen über den nähernden Ball. Es ist demnach möglich, auch in dem Fall der spontanen Handlung ein konkretes mentales Ereignis zu identifizieren, welches geeignet ist, als Relatum in der kausalen Beziehung zur durch es verursachten intentionalen Handlung zu fungieren. Ähnliches ließe sich für routinierte Handlungen durchführen: mein morgendliches Zähneputzen wird durch konkrete Überzeugungen – die Wahrnehmung des Weckerklingelns etc. – und durch sie aktualisierte dispositionale Wünsche verursacht. Problematisch an dieser Analyse ist sicherlich, dass sie den intuitiven Unterschied zwischen einer Handlung *aus* Gründen und Handlungen, für die eine Person Gründe hatte, sie zu tun, verwischt. Natürlich kann die Person, die den Ball fängt, Gründe angeben, weshalb sie den Ball gefangen hat. Diese rationalisieren ihre Handlung, sie erscheint vernünftig im Lichte ihrer bestehenden Überzeugungen, ihrer Wahrnehmung der Situation und dispositionale vorliegenden und in der Situation aktualisierten

---

<sup>157</sup> Vgl. Keil 2000, 144ff.

Wünschen. Die Person hat in diesen Fällen jedoch ihre Gründe, die sie für ihre Handlung hatte, nicht explizit für sich repräsentiert. Die Intentionalität ihrer Handlung war basal und hierin nicht wesentlich verschieden von der Intentionalität von Bewegungen, wie sie ein Hund, der einen Ball fängt, aufweist. Die Person kann im Unterschied zum Hund natürlich ihre Handlung ex post rationalisieren. Aber entscheidend ist, dass Personen auch vor ihren Handlungen die Gründe für verschiedene Handlungsoptionen repräsentieren und gegeneinander abwägen können. Die Person handelt nicht nur auf eine bestimmte Weise, weil bestimmte Gründe für diese Handlung sprechen, sondern *weil* sie außerdem einsieht, *dass* diese Gründe für diese Handlung sprechen. Der Unterschied ist analog zu dem Unterschied zwischen einer Regel bloß zu folgen und einer Regel zu folgen, *weil* man der Regel folgen will, oder einen richtigen Schluss zu ziehen und denselben Schluss zu ziehen, *weil* dieser aus den Prämissen logisch folgt. In Davidsons Theorie wird dieser Unterschied aus strukturellen Gründen verwischt.

Unterlassungen sind sowohl ontologisch – was *sind* Unterlassungen – als auch epistemologisch – wie *erkennen* wir Unterlassungen – eine Herausforderung für jede Handlungstheorie.<sup>158</sup> Ontologisch implizieren kausale Handlungstheorien das Stattfinden von Ereignissen wenn Handlungen vorliegen. Doch bestehen Unterlassungen gerade darin, dass etwas nicht geschieht. Das Nicht-Eintreten von Ereignissen selbst ist aber nur schwierig als Ereignis zu verstehen, und sogenannte ‚negative Ereignisse‘ anzunehmen ist auf den ersten Blick unattraktiv. Es bleibt noch, das Nicht-Geschehen im Falle von Unterlassungen als besondere Arten von Geschehnissen zu verstehen: es passiert doch etwas bei Unterlassungen, wir müssen nur genauer hinsehen, um es zu finden. Oder Unterlassungen werden nicht als Handlungen verstanden sondern als etwas anderes, wofür wir dennoch verantwortlich gemacht werden können. In jedem Fall provozieren Unterlassungen eigene Arten von Schwierigkeiten, die jedoch auch nicht-kausale Handlungstheorien betreffen und somit nur bedingt als spezifische Einwände gegen Davidsons kausale Handlungstheorie zählen.

### **e) Handlungen ohne Akteure**

Dieser Einwand besagt, dass in der kausalen Handlungstheorie Handlungen *ohne handelnde Personen* beschrieben werden. Der Anteil einer Person an ihren Handlungen bleibt vollkommen dunkel, und damit verfehlen kausale Handlungstheorien einen notwendigen Aspekt von Handlungen.

„I think that the standard story is flawed in several respects. The flaw that will concern me in this paper is that the story fails to include an agent – or, more precisely, fails to cast the agent in his proper role. In this story, reasons cause an intention, and an intention causes bodily movements, but nobody – that is, no person – *does* anything. Psychological and physiological events take place inside a person, but the person serves merely as the arena for these events: he takes no active part.” (Velleman 1992, 461)

---

<sup>158</sup> Für Unterlassungen vgl. Birnbacher 1995.

Diesem Einwand ist schwierig zu begegnen, da er komplizierte Fragen im Bereich personaler Identität, des Selbst u.a. aufwirft. *Erstens*: Die Adäquatheitsbedingung für die gelungene kausale Explikation des Handlungsbegriffs verlangt, dass die mentalen Ereignisse, welche die Handlung verursachen, oder das zusammengesetzte Ereignis, welches die Handlung ausmacht, nicht selbst wieder von solcher Art sind, dass sinnvoll die Frage gestellt werden kann, wie eine Person diese *vollziehen* kann. Die Explikation wäre gescheitert, wenn im Definiens auf Begriffe zurückgegriffen wird, die selbst nur über das Definiendum ‚(intentionale) Handlung‘ verständlich sind. In einem gewissen Sinne darf die Person innerhalb der kausalen Analyse *nichts tun*, was nicht selbst durch kausale Relationen zwischen Ereignissen analysierbar wäre.

*Zweitens* kann entgegnet werden, dass die Frage nach der Rolle der Person innerhalb der kausalen Analyse keinen rechten Sinn mache: nach dieser *handelt* die Person, wenn bestimmte mentale Ereignisse (Wünsche, Überzeugungen, Intentionen, Willensakte etc.) bestimmte andere Ereignisse (auf die richtige Weise) verursachen. Das ist das, was die Person *tut*.<sup>159</sup> Es ist somit zumindest sehr strittig, ob dieser Einwand überhaupt auf die kausale Handlungstheorie zutrifft und wenn ja, welchen Sinn dieser Einwand überhaupt macht unter der Annahme einer kausalen Handlungstheorie.

## **II.2.5 Probleme & Kritik II: Davidsons Konzeption von Willensschwäche**

Die Kritik an Davidsons Konzept der Willensschwäche im Speziellen kann unter anderem auf folgenden Ebenen ansetzen: die logische Verbindung der drei Thesen bzw. ihre unterstellte Widersprüchlichkeit; P3 und Davidsons Definition von Willensschwäche; die Prinzipien P2 und P3; Davidsons Unterscheidung zwischen prima facie Urteilen und unbedingten Urteilen sowie seine Teilungstheorie des Geistes.

### **a) Logische Widersprüchlichkeit von P1–P3**

Formal gesehen sind die drei Prinzipien in Gemeinschaft mit der Definition von Willensschwäche nur unter einer wohlwollenden Interpretation widersprüchlich. Schließlich ist genau der Umstand, den sich Davidson in seiner Auflösung des Trilemmas zunutze macht, ausschlaggebend für dessen (angebliche) logische Widersprüchlichkeit: In der Definition von Willensschwäche urteilt die Person alles in allem (all things considered), während sie in P2 ohne diesen Zusatz urteilt. Ohne die Annahme, diese beiden Formen des Urteilens seien logisch

---

<sup>159</sup> „Those who believe the story will of course contend that the events recounted in it add up to the agent’s participating in his action, as components add up to a composite. The story doesn’t mention his participation, they will explain, simply because his participation isn’t a component of itself. Complaining that the agent’s participation in his action isn’t mentioned in the story is, in their view, like complaining that a cake isn’t listed in its own recipe.” (Velleman 1992, 462)

miteinander verknüpft oder einander gleichwertig, sind die drei Prinzipien einander nicht widersprüchlich.

### **b) Prinzip P3 und Davidsons Definition von Willensschwäche**

These P3 besagt, dass es willensschwache Handlungen gibt. Diese Existenzbehauptung ist verknüpft mit einer spezifischen Definition von willensschwachen Handlungen, bei der die Adäquatheit der Definition in Frage gestellt werden kann. Die Existenzbehauptung selbst ist eine empirische Aussage: ob es letztlich willensschwache Handlungen gibt, ist eine Sache der Erfahrung, und lässt sich nicht aus rein begrifflichen Gründen zeigen. Die philosophische Analyse kann lediglich zeigen, dass Willensschwäche und willensschwache Handlungen aus begrifflichen Gründen *möglich* sind, indem sie eine verständliche Explikation anbietet, welche mit anderen grundlegenden Annahmen über die Welt zusammenstimmt. Ist der Begriff der Willensschwäche in sich oder im Zusammenhang mit anderen Überzeugungen inkonsistent, dann besteht ein guter Grund, entweder die Existenzbehauptung zu bestreiten, die Definition des Phänomens zu ändern oder die anderen Überzeugungen anzupassen, um Konsistenz zu erreichen.

An Davidsons Definition willensschwacher Handlungen kann kritisiert werden, dass sie zu eng sei: sie erfasse nicht alle Fälle von Willensschwäche, die wir aus anderen Gründen berechtigt seien, anzunehmen. Sie umfasse z.B. keine Fälle exekutiver Willensschwäche, in denen eine Person eine Intention für eine Handlung hat, und diese dennoch nicht ausführt. In Davidsons Verständnis von Intentionen als unbedingten evaluativen Urteilen wäre dies der Fall einer Person, die gegen ein unbedingtes evaluatives Urteil intentional handelt. Dieser Fall ist nach Davidsons Theorie ausgeschlossen.<sup>160</sup>

Die Definition kann auch als zu weit kritisiert werden: sie umfasse Handlungen als willensschwach, bei denen das bessere Urteil falsch ist oder auf irrationalen Wegen gewonnen wurde. Aristoteles kennzeichnet diese Fälle als uneigentliche aus, während bei Davidsons neutraler Definition diese normale Fälle willensschwachen Handelns darstellen. Auch kann kritisiert werden, dass sich Davidson alleine auf willensschwache Handlungen konzentriert, und das Phänomen der Willensschwäche als Charakterzug nicht weiter thematisiert. Schließlich kann kritisiert werden, dass Willensschwäche bei Davidson ein rein logischer Fehler sei: die Person *urteilt* nicht so, wie sie urteilen sollte gemäß dem Kontinenzprinzip. Zu dieser Kritik äußert sich Davidson folgendermaßen:

„However, I did not want to explain incontinence as like making a logical mistake. But the principle of continence is not a principle of logic. [...] I find it strange, though, to think of an incontinent intention or

---

<sup>160</sup> Zu diesem Einwand vgl. Mele 1987a; Pears 1984.

action as an error in belief, since I think of evaluative judgements as conative propositional attitudes. So to fail to reason to the right 'conclusion' means, in practical reasoning, to fail to form attitudes in a rational, coherent way. Among those attitudes are intentions. Failure to form an intention in accord with the principle of continence is, I still think, all too possible." (Davidson 1985a, 206)

Ob diese Erwiderung den Einwand entkräftet hängt davon ab, wie man Davidsons Unterscheidung zwischen dem *Fällen* eines Urteils und dem *Inhalt*, worüber geurteilt wird, versteht. Davidsons Prinzipien beziehen sich auf das Fällen (oder Aufrecht-Halten) von Urteilen, nicht auf den spezifischen Inhalt dieser Urteile. Auf diesen Punkt werde ich unten noch näher eingehen.

### **c) Die Prinzipien P1 und P2**

Davidson selbst sieht seine Thesen P1 und P2 als nicht optimal an, jedoch als ausreichend für die Charakterisierung des Problems, das Willensschwäche aufwirft.<sup>161</sup> P1 beschreibt einen Zusammenhang zwischen einer Handlung mehr erstreben, wünschen oder wollen und sie intentional tun, P2 zwischen evaluativen Urteilen über Handlungsoptionen und entsprechenden Wünschen (Volitionen, Strebungen), diese auch zu tun.

P1 hat starke Ähnlichkeit zu einem fundamentalen Grundsatz, der in vielen Handlungstheorien unterschiedlicher Prägung vorausgesetzt wird: Personen tun das, wozu sie am meisten motiviert sind. In der speziellen Form von Davidson ist dieser Satz eingeschränkt formuliert für intentionale Handlungen und nennt die ‚stärkere Motivation‘ nur als notwendige, nicht als hinreichende Bedingung für Handlungen. Nachdem Davidson intentionale Handlungen als Wirkung von bestimmten mentalen Ereignissen versteht, nämlich Kombinationen aus Überzeugungen und Wünschen (primären Gründen), ist unklar, wie sich diese kausale Analyse genau zu P1 verhält: schließlich referiert P1 nicht auf Überzeugungen der Person, sondern nur auf ihre Wünsche (Willen, Strebung). Oder anders: P1 formuliert keine kausal hinreichende Bedingung für das Stattfinden einer intentionalen Handlung, sondern bloß eine notwendige (if he does either *x* or *y* intentionally). Es heißt: ‚Wenn die Person intentional handelt, dann handelt sie gemäß ihres stärkeren Wunsches (Wille, Strebung)‘ und nicht ‚*dass* sie intentional handelt, wenn sie einen stärkeren Wunsch nach einer Handlung hat‘.

P2 hat ebenfalls Ähnlichkeit mit einem allgemeinen Grundsatz, wie er in Diskussionen um motivationalen Internalismus bezüglich moralischer oder allgemein evaluativer Urteile genannt

---

<sup>161</sup> „I am not very happy about P1–P3 as I have stated them: perhaps it is easy to doubt whether they are true in just their present form (particularly P1 and P2). And reflecting on the ambiguities, or plurality of uses, of various critical words or phrases (‘judge better’, ‘want’, ‘intentional’) it is not surprising that philosophers have tried interpreting some key phrase as meaning one thing in one principle and meaning something else in another. But I am convinced that no amount of tinkering with P1–P3 will eliminate the underlying problem: the problem will survive new wording, refinement, and elimination of ambiguity.” (Davidson 1969a, 23f.)

wird: Ein evaluatives Urteil impliziert notwendig die entsprechende Motivation der Person, diesem Urteil gemäß zu handeln.<sup>162</sup> Davidson versteht diesen Zusammenhang nicht als semantische These über die Bedeutung evaluativer Urteilsinhalte, sondern als Zusammenhang zwischen der Art und Weise, Urteile mit evaluativen Inhalten zu fällen, und der Motivation, diesen Urteilen entsprechend zu handeln. Hierin grenzt er sich von Präskriptivismus ab, wie er unter anderem von Richard Hare entwickelt und vertreten wurde.<sup>163</sup> Der Präskriptivismus ist eine semantische These über die Bedeutung von Ausdrücken wie ‚Ich soll X tun‘ oder ‚X ist gut zu tun‘. Diese Ausdrücke sollen präskriptiv verstanden werden, d.h. handlungsanleitend: ein aufrechter, kompetenter Sprecher verwendet diese Ausdrücke in ihrem eigentlichen (= präskriptiven) Sinne nur dann, wenn er auch motiviert ist, entsprechend dieser Urteile zu handeln. Dagegen wendet Davidson ein, dass dies den Unterschied zwischen dem *Inhalt* eines Urteils und dem *Modus*, in dem dieses gefällt wird, verwische:

„The confusion is between making a judgement, and the content of the judgement. It is P2 (or its ilk) that creates the problem, and P2 connects *making* a judgement with wanting and hence, *via* P1, with acting. But prescriptivism is a doctrine about the *content or meaning* of what is judged, and P2 says nothing about this.”  
(Davidson 1969a, 26)

Die Diskussion um motivationalen Internalismus bzw. seinem Gegenpart, dem motivationalen Externalismus, ist eines der Kernstücke der Diskussion um die Struktur und Beschaffenheit des Wissens der willensschwachen Person, das später noch näher geklärt werden muss.<sup>164</sup>

Auf andere Weise hat Gary Watson auf das dahinter liegende Problem hingewiesen, indem er bemerkt, dass sowohl P1 als auch P2 mehrdeutig seien, da ‚to want something more‘ zumindest auf zwei verschiedene Arten verstanden werden könnte: motivational und evaluativ.<sup>165</sup>

„To bring these points to bear on Davidson’s principles: let us distinguish two senses of “wants more” or “wants most,” an evaluational sense and motivational sense. In the first sense, if one wants to do *x* more than one wants to do *y*, one *prefers* *x* to *y* or ranks *x* higher than *y* on some scale of values or “desirability matrix.” In the second sense, if one wants to do *x* more than *y*, one is more strongly motivated to do *x* than to do *y*. Thus, *P2* may be true if understood in the language of evaluation, but false if understood in the language of motivation; whereas *P1* is true if understood in the language of motivation, but false if understood in the language of evaluation. [...] There is no univocal interpretation of the key phrases of *P1* and *P2* on which these principles turn out to be true, or even very plausible.” (Watson 1977, 320f.)

Die Unterscheidung zwischen einem motivationalen und evaluativen Sinne von ‚etwas mehr erstreben (wollen, wünschen)‘ ist natürlich eine bloße Neuformulierung des Problems, das Willensschwäche für den Zusammenhang zwischen dem Wissen einer Person und ihren Intentionen und Handlungen aufwirft. Auch wenn der Zusammenhang, den Davidson zwischen

---

<sup>162</sup> „Principle P2 states a mild form of internalism. It says that a judgement of value must be reflected in wants (or desires or motives).” (Davidson 1969a, 26) Vgl. Kap. III.5.

<sup>163</sup> Vgl. Kap. III.3.

<sup>164</sup> Vgl. Kap. III.5 und die Unterscheidung zwischen Akt- und Inhalts-Internalismus.

<sup>165</sup> Vgl. Taylor 1980; Frankena 1988.

Evaluation und Motivation in seinen Prinzipien entwirft, falsch sein möge, so bleibt das Problem bestehen, *wie* denn dieser Zusammenhang richtig zu verstehen ist.

Im Falle von Davidson impliziert seine kausale Handlungstheorie die beiden Prinzipien, und sogar deren Umkehrungen. Primäre Gründe verursachen Intentionen, die wiederum intentionale Handlungen verursachen und zugleich rechtfertigen. Intentionen wiederum analysiert Davidson als unbedingte evaluative Urteile, d.h. das *Fällen* solcher Urteile versteht Davidson als Bildung einer Intention, etwas zu tun. Damit sind Intentionen zugleich *inhaltlich evaluativ* – eine Handlung wird als erstrebenswert, gut, gesollt etc. beurteilt – und von ihrem *Modus* her *motivational* – das Fällen dieser Urteile ist *hinreichend* für intentionale Handlungen. Zu jeder intentionalen Handlung gibt es nach Davidson ein entsprechendes evaluatives Urteil, welche diese rationalisiert und verursacht – also die Umkehrung von P1. Zugleich heißt eine Handlung mehr tun zu wollen im Sinne eines konkreten Handlungsvorsatzes sie im Modus eines unbedingten evaluativen Urteils als erstrebenswert zu beurteilen – die Umkehrung von P2. Problematisch ist die Formulierung der beiden Prinzipien P1 und P2 aus dieser Sicht vor allem daher, da der Begriff des ‚etwas mehr wünschen (erstreben, wollen)‘ unterbestimmt bleibt ohne die kausale Analyse intentionaler Handlungen mit einzubeziehen. Dieses ‚etwas mehr wünschen‘ ist tatsächlich bei Davidson im Kontext dieser Analyse nicht irgendein Wunsch, sondern die Intention, eine Handlung auch durchzuführen. Aber, so könnte man einwenden, ist es nicht möglich, auch gegen vorliegende Intentionen, etwas zu tun, intentional zu handeln, d.h. wider ein unbedingtes evaluatives Urteil? Die Antwort auf diese Frage hängt damit zusammen, wie der Unterschied zwischen *prima facie* bzw. bedingten und unbedingten evaluativen Urteilen zu verstehen ist.

#### **d) Bedingte und unbedingte evaluative Urteile**

Wünsche und Strebungen lassen sich nach Davidson durch *prima facie* evaluative Urteile *ausdrücken*: Insofern die Handlung X die Eigenschaft Y besitzt, ist sie erstrebenswert. Intentionen, welche mit Handlungen einhergehen – ‚intensions in action‘ (Searle) bzw. handlungsleitende Intentionen – und Intentionen, die sich auf zukünftige Handlungen beziehen – ‚pure intendings‘ (Davidson) oder ‚future intentions‘ (Searle) – lassen sich durch unbedingte evaluative Urteile der Art ‚Handlung X ist erstrebenswert‘ ausdrücken. *Prima facie* evaluative Urteile fungieren als Gründe für Intentionen und die Handlungen, die sie verursachen. Bedingte evaluative Urteile über Handlungsoptionen werden im Hinblick auf ein beschränktes Feld an Überzeugungen und hinsichtlich bestimmter, eingeschränkter Eigenschaften von Handlungen getroffen. Ein Beispiel: Evi hat den Wunsch danach, eine kleine Katze zu streicheln. Das Streicheln der Katze ist damit aus ihrer Sicht *prima facie* erstrebenswert, nämlich insofern Evi das Streicheln des flauschigen Fells Freude bereitet. Aber Evi weiß auch, dass sie unter einer Katzenallergie leidet, sodass das

Streicheln der Katze zu einem unangenehmen Juckreiz und Ausschlag führen würde. Insofern ist das Streicheln der Katze aus ihrer Sicht *prima facie* nicht erstrebenswert. Dieselbe Handlung ist also sowohl erstrebenswert wie nicht erstrebenswert aus verschiedenen Perspektiven. Wenn nun Evi tatsächlich die Katze streichelt, dann kann die Beschreibung der Intention dieser Handlung nicht einfach aus den vorliegenden *prima facie* Urteilen, welche die verschiedenen Wünsche Evis bezüglich des Streichelns der Katze repräsentieren, abgeleitet werden, da diese einander widersprechen und Evi die Katze nicht gleichzeitig streicheln und nicht streicheln kann. Es braucht daher ein eigenes evaluatives Urteil, welches die Handlung nicht nur als erstrebenswert hinsichtlich einer ihrer aus einem bestimmten Grund erstrebenswerten Eigenschaften bewertet, sondern eines, dass die Handlung als unbedingt erstrebenswert auszeichnet.<sup>166</sup>

Die Frage ist, ob die Differenz zwischen ‚bloßen Wünschen oder Strebungen‘ und ‚Intentionen‘ tatsächlich verständlicher ist durch die Differenz zwischen *prima facie* und unbedingten evaluativen Urteilen, oder ob nicht vielmehr umgekehrt diese Differenz erst verständlich ist über den Begriff der Intention.

„There remains the question whether the sort of judgement to which I have referred, an all-out judgement, can be understood without appeal to the notions of intention or will. I asked at the beginning of this last section of my paper whether intending to do something is wanting to do it; if it were, we might consider that our aim had been achieved. What we intend to do we want, in some very broad sense of want, to do. But this does not mean that intending is a form of wanting. For consider the actions that I want to perform and that are consistent with what I believe. Among them are the actions I intend to perform, and many more. I want to go to London next week, but I do not intend to, not because I think I cannot, but because it would interfere with other things I want more. This suggests strongly that wanting and desiring are best viewed as corresponding to, or constituting, *prima facie* judgements. If this is correct, we cannot claim that we have made out a case for viewing intentions as something familiar, a kind of wanting, where we can distinguish the kind without having to use the concept of intention or will.” (Davidson 1978, 101f.)

Davidson gesteht demnach ein, dass letztlich die Differenz zwischen Intentionen und bloßen Wünschen oder Strebungen nicht über die Differenz zwischen unbedingten und bedingten evaluativen Urteilen verständlich wird. Es lassen sich nur beide durch solche evaluativen Urteile *ausdrücken*.<sup>167</sup> Dieser Umstand wirkt sich auf die Adäquatheit der kausalen Analyse des Handelns aus: Nicht nur abweichende Kausalketten, sondern auch der Begriff des unbedingten evaluativen Urteils, das als Bindeglied zwischen den primären Gründen und intentionalen Handlungen

---

<sup>166</sup> „When an intention is formed we go from a stage in which we perceive, or imagine that we perceive, the attractions and drawbacks of a course of action to a stage in which we commit ourselves to act. This may be just another pro-attitude, but an intention, unlike other desires or pro-attitudes, is not merely conditional or *prima facie*. If it is to produce an action, it can't be simply an appreciation that some good would come of acting in a certain way.” (Davidson 1987, 107)

<sup>167</sup> „What we can say, however, is that intending and wanting belong to the same genus of pro attitudes expressed by value judgements. Wants, desires, principles, prejudices, felt duties, and obligations provide reasons for actions and intentions, and are expressed by *prima facie* judgements; intentions and the judgements that go with intentional actions are distinguished by their all-out or unconditional form. Pure intendings constitute a subclass of the all-out judgements, those directed to future actions of the agent, and made in the light of his beliefs.” (Davidson 1978, 102)

fungiert, ist nicht verständlich ohne Rückgriff auf ein intuitives Vorverständnis dessen, was es heißt, eine Handlung zu intendieren.

### **e) Davidsons Teilungstheorie des Geistes**

Teilungsthesen des Geistes haben mit dem generellen Einwand zu kämpfen, bloße Problemverschiebung unter dem Deckmantel der Problemlösung zu leisten: das Problem, für das die Teilungsthese als Lösung ersonnen wurde, wiederholt sich einfach auf einer anderen Ebene, d.h. für die Teile selbst. So hat Jean-Paul Sartre in *Das Sein und Nichts* (1943) auf das Phänomen der Selbsttäuschung<sup>168</sup> bezogen gegen das Freud'sche Modell des Zensors argumentiert, welcher kontrolliert, welche Inhalte der Person bewusst werden und welche im Unbewussten oder Vorbewussten verbleiben: Ist sich der Zensor seiner Tätigkeit bewusst oder nicht? Ist er bewusst, so ist unklar, wie dieses Bewusstsein verschieden sein sollte vom ‚normalen‘ Bewusstsein der Person, ist er unbewusst ist fraglich, wie der Zensor seine Tätigkeit erfolgreich durchführen könnte.<sup>169</sup> Man könnte in Analogie ein ähnliches Argument im Falle von Willensschwäche bei Davidson formulieren: Welches Bewusstsein haben die mentalen Teile von der Irrationalität der Willensschwäche? Sind sich die Teile der Irrationalität bewusst, so wiederholt sich das Problem für die Teile selbst. Sind sie sich der Irrationalität nicht bewusst, dann kann die Irrationalität der Willensschwäche nicht als Verletzung eigener rationaler Standards verstanden werden, sondern nur als Verletzung fremder Standards. Oder anders: Willensschwäche ist dann wesentlich nur aus der Dritten-Person-Perspektive, nicht aus der Ersten-Person-Perspektive zuschreibbar. Weiters ist natürlich fraglich, wie die Rede der ‚Teilbarkeit des Geistes‘ genau verstanden werden soll. Was ist ein ‚mentaler Teil‘ und wie verhalten sich diese Teile zum Geist als Einheit, falls eine solche vorausgesetzt wird?<sup>170</sup>

### **Zusammenfassung**

Aristoteles beschäftigt sich mit dem Phänomen der Willensschwäche innerhalb seiner Tugendethik. Schon die bloße Existenz von Charaktertugenden zeugt davon, dass innerhalb dieser Theorie Wissen alleine noch nicht hinreichend ist für tugendhaftes Handeln, es braucht auch die richtige *Einstellung* zu den Wissensinhalten. Allgemein kommen bei Aristoteles Handlungen erst über das Zusammenspiel eines kognitiven und eines konativen Elementes

---

<sup>168</sup> Sartre diskutiert dieses Phänomen unter der Bezeichnung ‚Unaufrichtigkeit‘ (mauvaise foi).

<sup>169</sup> Vgl. Sartre 1943, 124ff. „Wenn wir also die Sprache und die verdinglichte [...] Mythologie der Psychoanalyse verwerfen, so bemerken wir, dass die Zensur kennen muss, was sie verdrängt, damit sie ihre Aktivität differenziert ausüben kann. [...] Kurz, wie könnte die Zensur die zu verdrängenden Triebe unterscheiden, ohne sich bewusst zu sein, dass sie sie unterscheidet? [...] Aber von welchem Typus kann das Bewusstsein (von) sich der Zensur sein? Es muss Bewusstsein (davon) sein, Bewusstsein des zu verdrängenden Triebs zu sein, aber gerade, *um nicht von ihm Bewusstsein zu sein*. Was kann das andres heißen, als dass die Zensur unaufrichtig sein muss?“ (Sartre 1943, 128)

<sup>170</sup> Vgl. Gardner 1993.

zustande. Dies führt zu einer Lesart der neuralgischen Kapitel der Nikomachischen Ethik zum Wissen des Akratikers welche einen deutlichen Abstand zwischen Aristoteles und Platons Sokrates einräumt. Der klarsichtige Willensschwache leidet demnach nicht an einem kognitiven Defizit – das sokratische Verständnis –, sondern er handelt aufgrund eines übermäßigen Affektes wider die als besser erkannte Handlungsoption.

Das Phänomen der Willensschwäche berührt zentrale Punkte innerhalb von Davidsons kausaler Analyse intentionalen Handelns. Diese expliziert Handlungen als kausale Wirkung bestimmter mentaler Ereignisse, welche über ihre propositionalen Gehalte die jeweilige Handlung rechtfertigen und als Ereignisse die Handlung zugleich verursachen. Das hauptsächliche Problem, das willensschwache Handlungen in diesem kausalen Rahmenmodell aufwerfen, ist die Unvereinbarkeit des intentionalen und irrationalen Charakters solcher Handlungen: sie sollen zugleich als irrational und intentional beschreibbar sein. Wenn aber intentionale Handlungen gemäß der kausalen Analyse verursacht werden durch Gründe, welche die Handlung auch rechtfertigen, dann ist fraglich, worin die Irrationalität der willensschwachen Handlung liegen soll. Davidson versteht die Irrationalität der Willensschwäche folglich als Meta-Irrationalität, als Inkohärenz zwischen einer Teilmenge an Gründen, die für die willensschwache Handlung sprechen, und einer diese Teilmenge umfassenden Menge an Gründen, welche für eine alternative Handlung sprechen, und nicht als Irrationalität zwischen der Teilmenge an Gründen und der durch sie verursachten willensschwachen Handlung. Die willensschwache Person intendiert nicht die Handlung, die sie auf Basis ihrer Gründe intendieren sollte. Sie verstößt damit gegen ein Prinzip der Anwendung praktischer Deliberation (das Kontinenz-Prinzip), die Intentionen nach den besten verfügbaren Gründen zu bilden bzw. nach den besten verfügbaren Gründen zu handeln. Doch weshalb verstößt der Willensschwache gegen dieses Prinzip, wie lässt sich Willensschwäche *erklären*? Hierfür greift Davidson auf das Konzept von Gründen, die bloße Ursache sind für andere mentale Ereignisse, ohne diese zu *begründen*, zurück. Im Grunde sind Davidsons handlungstheoretische Argumente rein negative Argumente gegen die These, dass Rationalisierungen nicht zugleich eine kausale Erklärung der involvierten Ereignisse sein können. Das positive Argument ist seine These, dass ohne eine kausale Interpretation unverständlich ist, wie man Fälle von echtem Handeln aus bestimmten Gründen von bloßen Rationalisierungen *ex post* und Fällen von Handlungen, in denen dem Akteur zugleich mehrere Gründe für eine Handlung zugänglich sind, dieser jedoch nur aufgrund einer Teilmenge dieser Gründe handelt, sonst unterscheiden sollte, wenn nicht über den Hinweis auf eine Verursachungsbeziehung. Wie diese kausale Beziehung zu interpretieren ist, dafür liefert Davidson kein Argument: er zeigt wiederum nur rein negativ, dass eine ereigniskausale und nomologische Deutung von Kausalität verträglich ist mit der Annahme, Rationalisierungen (oder intentionale Erklärungen) seien

Kausalerklärungen. *Das* aber Kausalität nur so verstanden werden kann, dafür gibt es kein eigenes Argument außer Hinweise auf David Hume und die traditionelle nomologische Deutung des Kausalbegriffs. Weiters gibt Davidson kein positives Argument dafür, weshalb Handlungen identisch sein sollen mit Ereignissen, genauer raumzeitlich lokalisierte Körperbewegungen, und weitergehend, weshalb Ereignisse überhaupt als ontologisch selbständige Entitäten aufzufassen sind. Wie sollen wir das Konzept von Ereignissen unabhängig von jeglicher Beschreibung verstehen? Diese allgemeinen Probleme die viele kausale Handlungstheorien betreffen, konnte ich hier nicht weiter verfolgen. Das Phänomen der Willensschwäche alleine liefert aufgrund ihres prekären Charakters keine hinreichenden Argumente für oder wider kausale Handlungstheorien. Abweichende Kausalketten und die Diskussion um ein adäquates Verständnis von Kausalität sind hierfür besser geeignet. Das Phänomen purer Intentionen hat außerdem gezeigt, dass Davidsons reduktive Analyse von Intentionen als besondere Formen von Urteilen (unbedingte evaluative Urteile) scheitert.

### III Problemfeld Evaluation und Motivation

In diesem Kapitel untersuche ich verschiedene Theorien, welche die Rolle des besseren Urteils für Willensschwäche untersuchen. Ich beginne mit der Diskussion von Selbsttäuschungstheorien von Willensschwäche bei Ursula Wolf und Julius Schälike (1), gefolgt von Theorien fehlerhafter Diskontierung von Platon und George Ainslie (2), dem Präskriptivismus von Richard Hare und deren Kritik durch Robert Dunn (3) sowie Theorien inkommensurabler Werte bei David Wiggins und deren Kritik durch Michael Stocker (4). Schließlich geht es grundlegend um den Zusammenhang zwischen Evaluation und Motivation, wie er im motivationalen Internalismus und Externalismus formuliert wird (5). Die Untersuchungen sind eher negativ, da sie Voraussetzungen und Argumente von Theorien in Frage stellen, welche in Konsequenz synchrone Willensschwäche als unmöglich erweisen. Dagegen werde ich eine Form des deskriptiven Externalismus verteidigen, der zwischen Evaluation und Motivation keinen strikten Zusammenhang ansetzt und so Raum lässt für diese Art der Willensschwäche.

#### III.1 Willensschwäche als Selbsttäuschung

Selbsttäuschungstheorien der Willensschwäche verstehen Willensschwäche als Selbsttäuschungen von Personen. Grundlegend unterscheiden sich Selbsttäuschungstheorien der Willensschwäche voneinander in der Bestimmung des genauen Gegenstandes, über den sich die Person selbst täuscht, und der Explikation der Möglichkeit einer solchen Selbsttäuschung. Ich werde Selbsttäuschungstheorien anhand von zwei Beispielen diskutieren, nämlich bei Ursula Wolf und Julius Schälike. Dabei soll deutlich werden, dass die Gründe, Willensschwäche *ausschließlich* als Selbsttäuschung zu verstehen, auf handlungstheoretischen und metaethischen Prämissen beruhen, die unabhängig von diesen Selbsttäuschungstheorien selbst sind.

##### III.1.1 Ursula Wolf: Zum Problem der Willensschwäche

Ausgehend von der klassisch-antiken Diskussion bei Sokrates und Aristoteles führt Ursula Wolf das Phänomen der Willensschwäche im Kern als das Phänomen ein, „dass jemand nicht tut, was er für das Beste hält, obwohl er es tun könnte.“ (Wolf 1985, 21) Diese Beschreibung des Phänomens sei jedoch nach Wolf in sich widersprüchlich, da das Ergebnis einer Überlegung darüber, was für mich in einer konkreten Situation am besten zu tun ist, die Form eines Handlungsvorsatzes hat, und Handlungsvorsätze per definitionem ausgeführt werden, wenn ich an ihrer Ausführung nicht gehindert bin.<sup>171</sup>

---

<sup>171</sup> Vgl. auch zuvor in ihrer zustimmenden Interpretation der aristotelischen Position: „Die praktische Überlegung fragt nach der hier und jetzt und für mich besten Handlung, gegeben meine verschiedenen handlungsbezogenen Wünsche, meine Fähigkeiten sowie die Situationsumstände, und ihr Ergebnis hat da her unmittelbar die Form eines

„In der praktischen Überlegung geht es nicht um irgendein Bestes, sondern um die Frage, was hier und jetzt für mich zu tun das Beste ist, und daher hat die Antwort die Form eines Handlungsvorsatzes. Da wir jetzt davon ausgehen, dass der Handelnde in der Situation so oder anders handeln kann, folgt aus dem Vorliegen eines Vorsatzes analytisch die Ausführung.“ (Wolf 1985, 26f.)

Das Argument lässt sich folgendermaßen verstehen:

- (P1) Wenn eine Person eine praktische Überlegung anstellt, dann überlegt sie darüber, was für sie selbst in einer konkreten Situation am besten zu tun ist.
- (P2) Wenn eine Person überlegt, was für sie selbst in einer konkreten Situation am besten zu tun ist, dann ist das Ergebnis dieser Überlegung ein Handlungsvorsatz.
- (K1) Also: Wenn eine Person eine praktische Überlegung anstellt, dann ist das Ergebnis dieser Überlegung ein Handlungsvorsatz.
- (P3) Wenn eine Person einen Handlungsvorsatz hat, dann führt sie diesen entweder aus, oder sie kann in der Situation nicht so oder anders handeln.
- (K2) Wenn eine Person eine praktische Überlegung anstellt und zu einem Ergebnis gelangt, dann handelt sie entweder entsprechend der Überlegung, oder sie kann nicht so oder anders handeln.

Diese Überlegung führt zu einem Widerspruch mit der folgenden Beschreibung von Willensschwäche:<sup>172</sup>

- (P1) X kann so oder anders handeln.
- (P2) X stellt eine praktische Überlegung an und kommt zu einem Ergebnis.
- (P3) X handelt gegen ihre praktische Überlegung.

Da Wolf die Prämissen ihres Argumentes akzeptiert, sollten wir ihrer Meinung nach diese Beschreibung des Phänomens ablehnen und uns auf die Suche nach einer nicht-widersprüchlichen Beschreibung machen. Passt man die Beschreibung von Willensschwäche an, so lässt sich der Widerspruch auf folgende Weisen auflösen: a) Eine willensschwache Person kann nicht so oder anders handeln;<sup>173</sup> b) Eine willensschwache Person stellt keine *praktische* Überlegung an oder die willensschwache Person gelangt nicht zu einem Ergebnis in ihrer

---

Handlungsvorsatzes. Ein konkreter Handlungsvorsatz hinsichtlich einer jetzt anstehenden Handlungssituation ist aber dadurch definiert, dass er zur Ausführung kommt, wenn kein äußeres Hindernis vorliegt und der Handelnde ihn ausführen kann [...]. Wenn daher jemand den Vorsatz, die nach seiner Überlegung beste Handlung zu tun, nicht ausführt, obwohl kein äußeres Hindernis vorliegt und obwohl er die Fähigkeit im gewöhnlichen Sinn der physischen usw. Fähigkeit besitzt, dann folgt analytisch, dass er entweder den Vorsatz nicht wirklich hat oder ihn aufgrund bestimmter innerer Bedingungen nicht ausführen kann.“ (Wolf 1985, 22)

<sup>172</sup> „Wenn die Meinung über das Beste Vorsatzcharakter hat, dann verwickeln wir uns in einen Widerspruch, wenn wir sagen, dass jemand das, was er für das in der Situation Beste hält, in dieser Situation tun konnte, es aber doch nicht tut. Ich stimme daher Sokrates und Aristoteles zu, dass es dieses Phänomen in der Tat nicht geben kann.“ (Wolf 1985, 23) Ob Sokrates und Aristoteles tatsächlich so zu verstehen sind diskutiere ich in Kap. III.2.1 und II.1.

<sup>173</sup> Nach Wolf ist dies die Position von Richard Hare. Vgl. Wolf 1985, 24; Kap. III.3.

praktischen Überlegung; c) Eine (vermeintlich) willensschwache Person handelt nicht gegen ihre praktische Überlegung, sondern täuscht sich selbst.

Nach Wolf ist Option a) unattraktiv, da revisionär gegenüber unserer Praxis. Außerdem führt sie ein Argument von Anthony Kenny an,<sup>174</sup> das gegen Hares Auffassung gerichtet ist, die willensschwache Person sei psychisch unfähig, anders zu handeln, als sie es tut:

„Die praktische Überlegung fragt nach der besten der mir möglichen Handlungen, und daher kann, wenn ich mich selbst so sehe, dass ich unter innerem Zwang stehe und nicht anders handeln kann, die betreffende Handlung nicht Inhalt meiner Meinung über das Beste sein.“ (Wolf 1985, 25)

Dieses Argument setzt voraus, dass die willensschwache Person sich selbst so sieht, dass sie unter innerem Zwang steht und daher nicht anders handeln könne. Es wäre jedoch möglich, dass eine Person fälschlich glaubt, so oder anders handeln zu können, obwohl sie es tatsächlich nicht kann. In diesem Falle würde die Person sich darüber täuschen, welche Handlungsalternativen ihr tatsächlich offenstehen, und könnte dennoch eine praktische Überlegung im Sinne Wolfs anstellen und gegen diese handeln.<sup>175</sup> Im Grunde wäre dies eine Version der Explikation von Willensschwäche als Selbsttäuschung, wobei die Person sich täuscht, dass sie anders handeln könnte. Laut Wolf liegt jedoch das Phänomen der Willensschwäche zwischen zwanghaftem Handeln und dem Handeln aus den besten Gründen, weshalb wir versuchen sollten, eine Beschreibung von Willensschwäche zu finden, die kompatibel ist mit dieser Auffassung.<sup>176</sup>

Option b) diskutiert Wolf anhand von Donald Davidson und Irving Thalberg.<sup>177</sup> Aus ihrer Sicht beschreiben sie die willensschwache Person so, dass diese praktisch überlege und zu einem Ergebnis gelange, das Ergebnis aber aufgrund eines starken Wunsches sofort wieder revidiere. Die Person handelt aus einem Grund hinsichtlich des starken Wunsches, aber aus keinem Grund hinsichtlich des Umstandes, gegen die als beste erachteten Gründe zu handeln. Laut Wolf ist diese Beschreibung widersprüchlich, da die Person nicht zugleich eine praktische Überlegung anstellen, zu einem Ergebnis gelangen, und gegen dieses aus freien Stücken handeln könne. Dies deshalb, da wie oben beschrieben das Ergebnis einer praktischen Überlegung ein Handlungsvorsatz sei. Wolf lehnt deswegen diese Beschreibung von Willensschwäche ab. Stattdessen müsste diese so beschrieben werden, dass die Person entweder nicht praktisch überlegt oder in ihrer praktischen Überlegung zu keinem Ergebnis gelangt ist. Das Problem an dieser Option ist, wie die Person gegen eine praktische Überlegung handeln sollte, die sie gar nicht anstellte oder in der sie zu gar keinem Ergebnis gelangte.

---

<sup>174</sup> Vgl. Kenny 1975.

<sup>175</sup> Ob dies jedoch tatsächlich noch als Handlung anzusehen ist, ist fraglich.

<sup>176</sup> Die Frage der Abgrenzung von zwanghaftem und willensschwachem Handeln werde ich später noch eingehender diskutieren. Vgl. Kap. V.3.

<sup>177</sup> Vgl. Wolf 1985, 26f.; Davidson 1969a; Thalberg 1971.

Gegen diese *reductio* lässt sich jedoch einwenden, dass Davidson den Begriff der praktischen Überlegung differenzierter als Wolf versteht, sodass er mithin nicht die Beschreibung von Willensschwäche angreift, sondern eine der Prämissen von Wolfs Argument, nämlich Prämisse 2 – Wenn eine Person überlegt, was für sie selbst in einer konkreten Situation am besten zu tun ist, dann ist das Ergebnis dieser Überlegung ein Handlungsvorsatz. Die willensschwache Person nach Davidson gelangt nur zu einem *prima facie* Urteil in ihrer praktischen Überlegung, und nicht zu einem *unbedingten* Urteil. Das unbedingte Urteil versteht er wie Wolf als Handlungsvorsatz oder Intention, das *prima facie* Urteil ist jedoch keine Intention, dennoch dasjenige, was aus Sicht der willensschwachen Person das *bessere* Urteil ist bzw. das Urteil über das in dieser Situation Beste. So kann die willensschwache Person laut Davidson wider das bessere Urteil und dennoch intentional handeln. Sie kann sich für die Handlungsoption entscheiden, die nicht dem besten Urteil entspricht. Bei Wolf bedeutet hingegen ‚das bessere Urteil‘ oder ‚etwas für das Beste halten‘ notwendigerweise, dieses Beste auch zu wollen bzw. entschieden anzustreben, mithin einen Handlungsvorsatz zu bilden.

„Wenn jemand sagt, dass er etwas für das Beste hält, aber doch anders handelt, können wir also in der Tat nur noch die vorhin schon erwähnte Konsequenz ziehen, dass er es nicht wirklich für das Beste hält bzw. nicht wirklich im handlungsrelevanten Sinn will.“ (Wolf 1985, 28)

Nach Davidson sollten wir jedoch Urteile *über etwas Praktisches* – evaluative Urteile über Handlungsoptionen – von genuin praktischen Urteilen – Handlungsintentionen verstanden als unbedingte evaluative Urteile – unterscheiden. Mit anderen Worten: Handlungsvorsätze sind weder identisch mit noch folgen diese logisch aus Urteilen darüber, was in einer Situation am besten zu tun ist. Eine Person kann demnach praktisch überlegen, zu einem Ergebnis gelangen, und dennoch entgegen dieser Überlegung handeln.

Wolf selbst präferiert Option c). Wolf gelangt dazu über eine Diskussion des Vorschlags von Anthony Kenny, zwischen unmittelbaren und höherstufigen Wünschen zu unterscheiden.<sup>178</sup> Ein höherstufiger Wunsch ist ein Wunsch, der als Objekt einen anderen Wunsch hat. Die (vermeintlich) willensschwache Person handle aufgrund eines unmittelbaren Wunsches, der aus Sicht der Person in der Situation besser sei als der höherstufige Wunsch. Deshalb verlöre aber der höherstufige Wunsch nicht gleich jede Handlungsrelevanz, wenn er nicht in jeder Situation realisiert wird. Dennoch meint Wolf, dass der höherstufige Wunsch zumindest an Handlungsrelevanz *einbüße*, sobald er nicht in der entsprechenden Situation realisiert wird. Problematisch an dieser Beschreibung ist, dass man das Handeln der willensschwachen Person so aber nicht als Handeln wider das in einer konkreten Situation beste Urteil beschreiben kann: Die

---

<sup>178</sup> Vgl. Wolf 1985, 27f.; Kenny 1975; Frankfurt 1971.

willensschwache Person handelt nicht gegen ihre praktische Überlegung, da sie genau das tut, was sie in Wahrheit für das Beste in dieser Situation hält.

„Wer gegen seine Meinung über das Beste seinem unmittelbaren Wollen folgt, müsste also sehen, dass er damit in Wirklichkeit seine Meinung geändert hat und damit seine Überlegungen revidieren müsste.“ (Wolf 1985, 28)

Doch worin besteht Willensschwäche dann nach Wolf? In der Selbsttäuschung der willensschwachen Person über die Handlungsrelevanz ihrer höherstufigen Wünsche. Die willensschwache Person glaube nämlich weiterhin, dass, auch wenn sie in ihrem Handeln ihren unmittelbaren Wünschen folgt, ihre höherstufigen Wünsche wirklich handlungsrelevant seien. Und das Motiv, weshalb wir dies tun, ist, dass wir uns gerne als eine Person mit gewissen handlungsrelevanten (höherstufigen) Wünschen vorstellen und mit einer solchen identifizieren:

„Man könnte daher sagen, dass wir es hier letztlich mit einem kognitiven Fehler zu tun haben, wenn auch von besonderer Art, nämlich einem Fehler im Selbstverständnis. Dieser Fehler ist ein interessierter Fehler. Wir machen ihn deswegen, weil wir einerseits vor uns selbst und vor anderen jemand sein möchten, der bestimmte höherstufige Wünsche als wirkliche Handlungsabsichten hat, weil wir aber andererseits im Konfliktfall ungern auf die Befriedigung unmittelbarer Wünsche verzichten.“ (Wolf 1985, 29)

Die Person irrt sich damit nicht nur über ihre handlungsrelevanten Wünsche, sondern sie täuscht sich selbst. Wolf skizziert dabei zwei Verfahren deren sich eine Person bedienen kann um sich selbst zu täuschen: Rationalisierungen durch Ausnutzung von Überlegungsspielräumen und Überhöhungen von unmittelbaren Wünschen sowie das Ausnutzen von Zeitdifferenzen zwischen dem Fassen von Vorsätzen und der Ausführung von Handlungen.<sup>179</sup>

Wolfs Beschreibung von (vermeintlicher) Willensschwäche als Selbsttäuschung ist problematisch. Zuerst stellt sich die Frage, worüber genau sich die willensschwache Person selbst täuscht. Täuscht sie sich darüber, was sie in einer konkreten Situation für das Beste hält? Täuscht sie sich darüber, was sie wirklich will? Nach Wolf überlegt die willensschwache Person, was sie in einer konkreten Situation am besten tun sollte. Das Ergebnis einer solchen Überlegung ist ein Handlungsvorsatz. Dieser wird ausgeführt, solange die Person ihn ausführen kann und den Handlungsvorsatz nicht ändert. Nach Wolf fallen demnach Handlungsvorsatz und die Meinung über das Beste zusammen. Wenn sich eine Person nun über das täuschte, was sie aktuell für das Beste halte, dann täuschte sie sich über ihren eigenen Handlungsvorsatz. Dies hieße, dass eine Person einen Handlungsvorsatz ausführte ohne zu wissen worin dieser bestünde. Damit würde die willensschwache Handlung aber nicht mehr dem Kriterium der Wissentlichkeit genügen, das Wolf selbst in ihrer Diskussion um Willensschwäche als Zwangshandlung anwendet. Die Person täuscht sich demnach nicht *aktuell* über ihre Meinung über das Beste, sondern entweder vorher oder nachher: Vorher im Sinne der Rationalisierung durch Überhöhung von unmittelbaren

---

<sup>179</sup> Für diese Strategien vgl. die Diskussion in Kap. III.1.2 und Wolf 1985, 30ff;

Handlungswünschen – Wolf spricht hier von Selbstdistanzierungen – und nachher im Sinne der Selbstzuschreibung von Gründen für die willensschwache Handlung, aus denen die Handlung aber in Wirklichkeit nicht getan wurde. Außerdem besteht die Möglichkeit, dass zwischen allgemeinen und konkreten Handlungsvorsätzen – unmittelbares und höherstufiges Wollen – Überlegungsspielräume ausgenutzt werden. Willensschwäche ist damit ein *diachrones* Phänomen, das sich im Wechsel zwischen Meinungen darüber, was eine Person am besten tun sollte, ausdrückt, wobei die Person sich lieber mit der einen Meinung identifiziert und sich diese zuschreibt als mit der anderen. Außerdem ist Willensschwäche hier nur als exekutive Willensschwäche im Blick, als Problem der Umsetzung allgemeiner Handlungsvorsätze oder dem kurzfristigen Wechsel von konkreten Handlungsvorsätzen. Motivationale Willensschwäche, als Handeln wider das bessere gegenwärtige *Urteil*, wird hingegen bereits aufgrund der handlungstheoretischen Voraussetzungen ausgeschlossen.

Bei ihrer Ablehnung von synchronen Formen von Willensschwäche stützt sich Wolf auf die Gleichsetzung von Meinungen über das Beste – als Ergebnisse von praktischen Überlegungen – und Handlungsvorsätzen. Für diese Gleichsetzung wird jedoch nicht argumentiert. Weshalb sollte die Meinung darüber, was ich jetzt und hier am besten tun sollte, *dasselbe* sein wie ein Handlungsvorsatz? Und umgekehrt: Warum sollte das, was ich vorsätzlich tue, *immer* dem entsprechen, was ich für das Beste zu tun halte? Für beide Thesen argumentiert Wolf nicht, außer einem Hinweis auf ihre präskriptivistische Interpretation von Wertsätzen über die hier und jetzt beste Handlung.<sup>180</sup> Dies, obwohl diese Frage den eigentlichen Kern der Debatte um die Möglichkeit von synchroner, motivationaler Willensschwäche berührt. In Frage steht dabei auch die Auffassung, was eine praktische Überlegung ist und was sie leistet. Welche Frage versucht eine Person durch eine ‚praktische Überlegung‘ zu lösen und weshalb ist diese Überlegung ‚praktisch‘?<sup>181</sup> Versucht sie herauszufinden, *was* sie tun sollte, oder was *am besten* zu tun ist? Und ist das Ergebnis einer praktischen Überlegung ein Handlungsvorsatz und daher ‚praktisch‘, oder ein abschließendes Urteil über etwas Praktisches (die eigene Handlung), die dann ‚praktisch‘ wird, indem die Person sie tatsächlich intendiert umzusetzen?

Weiters ist problematisch, dass eine Explikation von Willensschwäche als Selbsttäuschung ohne Explikation von Selbsttäuschung die philosophischen Probleme nur verschiebt. Das Phänomen

---

<sup>180</sup> „Was diese Wertsätze betrifft, mit denen ich nach der hier und jetzt besten Handlung frage, würde ich also Hare rechtgeben, dass sie „präskriptiv“ sind. Bei allen anderen Arten von Wertsätzen scheint mir seine pauschale Präskriptivitätsthese hingegen unsinnig. Da die Wertsätze, die nach der hier und jetzt für mich besten Handlung fragen, trivialerweise präskriptiv sind, ist derjenige Diskussionsstrang in der Debatte um die Willensschwäche, der sich an der Frage orientiert, was es heißt, einem praktischen Prinzip zuzustimmen oder eine praktische Meinung zu haben, uninteressant.“ (Wolf 1985, 27) Hares These diskutiere ich näher in Kap. III.3.

<sup>181</sup> Diese Frage verfolge ich näher in Kap. VI.1

der Selbsttäuschung ist schließlich mindestens ebenso umstritten wie das Phänomen der Willensschwäche, und es bedarf einiger philosophischer Anstrengung um die vielen verwickelten Probleme im Zusammenhang mit Selbsttäuschung zu lösen.<sup>182</sup> Wolf nennt zwar Strategien von Selbsttäuschung, doch ist unklar, wie die Person diese genau umsetzt.<sup>183</sup> Wie kann die Person zugleich glauben, dass sie ihre Meinung über das Beste nicht geändert habe, wenn sie doch gemäß einer davon abweichenden Meinung gehandelt hat und uns unsere eigenen Handlungsvorsätze unmittelbar zugänglich sind?

### III.1.2 Julius Schälike: Willensschwäche als Selbsttäuschung

In seinem Aufsatz *Willensschwäche und Selbsttäuschung* (2004) führt Schälike das Phänomen der Willensschwäche ein als das Phänomen, dass „die Person nicht das tut, was sie für das Beste hält, obwohl sie es tun könnte.“ (Schälike 2004, 361) Im Kern geht es auch hier um die Frage, ob synchrone motivationale Willensschwäche möglich ist oder nicht. Das philosophische Problem, das diese Beschreibung von Willensschwäche als Handeln wider das bessere Wissen aufwirft, skizziert er so:

„Unternimmt man jedoch den Versuch, Phänomenbeschreibungen von Willensschwäche in Einklang mit fundamentalen Prinzipien einer plausiblen Handlungstheorie zu bringen, so erweist sich dies als überraschend schwierig. Das liegt an der Annahme, dass willensschwache Handlungen *absichtliche* Handlungen sind. Absichtliches Handeln aber ist *frei* in dem Sinne, dass es einen Spielraum gibt, innerhalb dessen es von der Entscheidung der Person abhängt, ob sie so oder anders handelt, dass es also *an ihr selbst* liegt, was sie tut. Wer Handlungsfreiheit in diesem Sinne genießt, der *muß* irgendwann abschließend Stellung zu seinen Optionen beziehen (denn wählen zu können bedeutet, wählen zu müssen), und was dann geschieht, wird die Folge seiner Stellungnahme, seiner „Entscheidung“ sein. Versteht man diese Stellungnahme als Urteil, was zu tun situativ unter Berücksichtigung aller Umstände am besten ist, so ist „Willensschwäche“ ein paradoxer Begriff, das Phänomen somit inexistent.“ (Schälike 2004, 362)

Willensschwäche als das Phänomen, das eine Person wider ihr bestes praktisches Urteil frei und intentional handelt, führt zu Paradoxien innerhalb von Handlungstheorien, welche das beste Urteil als Intention interpretieren.<sup>184</sup> Schälike unterscheidet dabei zwischen Kongruenz- und Diskrepanzmodellen bzw. zwischen monistischen und dualistischen Modellen von Handlungstheorien. Kongruenzmodelle gehen von einer Identität von bestimmten Werturteilen und Intentionen aus, während Diskrepanzmodelle diese Identität verneinen. Die praktische Überlegung ist nach dem Kongruenzmodell instrumentell eingebettet in das Handlungsvermögen und deren Ergebnisse sind logisch notwendig mit praktischen Stellungnahmen verbunden: Eine Handlung unter verschiedenen Handlungsalternativen als beste zu beurteilen heißt innerhalb

---

<sup>182</sup> Für einen Überblick vgl. Mele 1987c.

<sup>183</sup> Dabei ist sich Wolf bewusst, dass ihre Untersuchung der Willensschwäche eine Untersuchung der Selbsttäuschung erforderlich macht. Vgl. Wolf 1985, 33.

<sup>184</sup> „Nimmt man einen analytischen Zusammenhang zwischen den Begriffen „Handlungsfreiheit“, „Entscheidung“ und situatives, abschließendes handlungsbezogenes Werturteil“ an, so ist der Begriff eines Handelns wider das bessere situative Urteil widersprüchlich.“ (Schälike 2005, 23)

dieser Modelle, diese Handlung auch zu wollen bzw. sich für diese Handlungsalternative zu entscheiden. Dualistische Modelle hingegen sehen die praktische Überlegung als relativ autonom gegenüber dem Handlungsvermögen an: es ist eine Sache, Werturteile über Handlungsalternativen zu fällen und eine bestimmte Handlung als beste zu beurteilen, und eine andere, diese auch tatsächlich zu wollen bzw. sich für die Umsetzung dieser Alternative zu entscheiden.<sup>185</sup> Schälike vertritt eine monistische Konzeption und kritisiert dualistische Konzeptionen hinsichtlich des Phänomens der Willensschwäche aus folgenden Gründen:

*Erstens*<sup>186</sup> sei unklar, was eine willensschwache Person tun könne, um ihre Willensschwäche zu überwinden. Eine erhöhte Konzentration auf die Gründe, die zu dem abschließenden besseren Urteil gehören, reiche nicht aus:

„Man entscheidet irrational nicht deshalb, weil man nicht *deutlich genug* erkannt hat, was zu tun wirklich am besten wäre, also unabsichtlich, sondern weil man irrational sein *will*, und das heißt, weil der Charakter auf der Entscheidungsebene irrational disponiert ist.“ (Schälike 2004, 367)

Hier vermischt Schälike jedoch die Intentionalität der willensschwachen Handlung mit dem Bewusstsein der Person um ihre Willensschwäche: eine dualistische Theorie muss nicht davon ausgehen, dass die willensschwache Person *willensschwach sein will*, wenn sie willensschwach handelt. Die willensschwache Person könnte sich für die willensschwache Handlung entscheiden ohne die Absicht zu haben, dabei willensschwach zu sein. Ob diese Möglichkeit konsistent vertreten werden kann, hängt vor allem davon ab, wie man die Intentionalität von Handlungen versteht und was es heißt, etwas intentional zu tun. Neben der erhöhten Konzentration auf die Handlungsgründe helfe nach Schälike das planmäßige Vermeiden von Situationen, in denen willensschwache Entscheidungen drohen, auch nicht, um Willensschwäche gemäß einer dualistischen Theorie zu vermeiden. Die planmäßige Vermeidung kann schließlich aus Willensschwäche unterlassen werden, was einen Teil von willensschwachen Handlungen unanalysiert ließe. Schließlich helfe auch Erziehung oder Konditionierung nicht, um die Willensschwäche zu beseitigen, da diese beiden Methoden nur auf die praktische Überlegung Einfluss nehmen und nicht auf die Tugend des Entscheidens.<sup>187</sup> Dass Erziehung und Konditionierung nur über den Umweg der praktischen Überlegung auf das Handeln einwirkt, setzt jedoch bereits das monistische Konzept voraus um plausibel zu sein. Weshalb sollte Erziehung darauf beschränkt

---

<sup>185</sup> Vgl. zur Gegenüberstellung dieser beiden Modelle Schälike 2004, 364ff. und Schälike 2005, 22ff.

<sup>186</sup> Vgl. Schälike 2004, 366f.

<sup>187</sup> „Erziehung, so scheint mir, kann jedoch nur einen Einfluss auf die Wünsche und die Überlegungsfähigkeit ausüben, nicht aber auf die Disposition, rational oder irrational zu entscheiden. Wie sollte die Disposition, auf die Vernunft, also auf Argumente zu hören, durch Argumente beeinflusst werden?“ (Schälike 2004, 368) In diesem Zusammenhang erwähnt Schälike auch Aristoteles: Charaktertugenden trügen nur zur korrekten praktischen Überlegung bei. Ich bin hier anderer Ansicht: Charaktertugenden haben ihr autonomes Gebiet, nämlich Strebungen und Affekte, unabhängig davon, wie gut oder schlecht eine Person praktisch überlegt. Vgl. Kap. II.1.

sein, lediglich auf die praktische Überlegung, und nicht auch direkt auf die Strebungen und affektiven Tendenzen einer Person einzuwirken?

*Zweitens*<sup>188</sup> seien unsere intentionalen Handlungen die primären Indikatoren für die Zuschreibung evaluativer Einstellungen. Unsere Praxis schliesse den Fall sogar aus, dass wir einer Person evaluative Einstellungen zuschrieben, die sie niemals ins Handeln umsetzte, wenn sich ihr die Gelegenheit dazu böte und sie entsprechend handeln könnte.

„Wir halten es nicht nur für sehr viel *wahrscheinlicher*, dass jemand, der sich niemals im Geringsten beim Essen zurückhält, obwohl er behauptet, unbedingt abnehmen zu wollen, lügt oder sich selbst täuscht, als dass er sich immer wieder seinem Urteil zuwider entscheidet, sondern wir halten uns für berechtigt, die Möglichkeit, dass er entsprechende evaluative Einstellungen hat, definitiv auszuschließen.“ (Schälike 2004, 369)

Dualistische Konzeptionen können natürlich zustimmen, dass solche Fälle äußerst selten sind, doch können sie darauf verweisen, dass die eigentliche Frage nicht die ist, ob wir solchen Personen diese evaluativen Einstellungen zuschreiben oder nicht, sondern ob diese Personen diese evaluativen Einstellungen haben könnten, obwohl sie ständig etwas anderes tun. Diese Frage betrifft schwierige Fragen in der Philosophie des Geistes. Wer meint, dass es geistige Einstellungen gibt, die sich nicht über bestimmte Handlungsdispositionen definieren lassen, der muss die theoretische Möglichkeit einräumen, dass eine Person solche evaluativen Einstellungen haben könnte, auch wenn sie niemals entsprechend handelt. Diese Frage ist wiederum von der Frage verschieden, ob eine Person *global* gesehen immer entgegen ihrer besseren Urteile handeln könnte, mithin ob globale Willensschwäche eine theoretische Option darstellt.<sup>189</sup> Aber auch hier könnte eine dualistische Konzeption die Möglichkeit globaler Willensschwäche verneinen, ohne damit die Möglichkeit vereinzelter, auch radikaler, Inkongruenzen zwischen evaluativen Einstellungen und Handlungen gleich mit zu verneinen. Für die Frage, ob wir der konkreten Person die evaluative Einstellung – z.B. ‚Abnehmen wäre für mich alles in allem die beste Handlungsoption‘ – zuschreiben oder nicht, ist vor allem entscheidend, welcher Art der Konflikt ist, den wir der Person zuschreiben, bzw. für wie kompliziert wir diese Person halten. Schreiben wir der Person Selbsttäuschung zu, dann ist der Konflikt ein anderer als wenn wir ihr Willensschwäche attestieren. In beiden Fällen verlangt offenbar das Verhalten der Person nach einer Deutung, die komplexer ist als das bloße Zuschreiben von evaluativen Einstellungen und entsprechenden Handlungen, da sie widersprüchliche Evidenzen liefert, die ihr Handeln und ihre Wertüberzeugungen als teilweise inkohärent zueinander erscheinen lassen. Auch kann eingewendet werden, dass die Kriterien für die Zuschreibung evaluativer Haltungen vielfältiger sind als nur das Erfüllungskriterium in entsprechenden Handlungen. Charakteristischerweise wird

---

<sup>188</sup> Vgl. Schälike 2004, 369f.

<sup>189</sup> Vgl. Kap. VI.5.

die willensschwache Person sowohl Gründe liefern, ihr die evaluative Haltung zuzuschreiben, als auch solche, sie ihr abzusprechen.

*Drittens*<sup>190</sup> sei es vor dem Hintergrund einer dualistischen Konzeption merkwürdig, dass willensschwache Handlungen immer dann auftreten, wenn Personen vor der Wahl zwischen Handlungen mit raschem Lustgewinn und solchen, die mit hohem Sozialprestige, aber wenig Lustgewinn verbunden sind, stehen. Weshalb entscheidet sich die willensschwache Person für das Essen der Schokoladentorte, wenn sie das Abnehmen für besser hält, und nicht für das Wegwerfen der Schokoladentorte, wenn sie das Essen für besser hält? Die einfache Antwort auf diese Frage wäre, dass das bessere Urteil inhaltlich meist gegen den raschen Lustgewinn und für den Lustaufschub spricht, dass mithin die allermeisten keinen zügellosen Hedonismus befürworten. Für diese These bedarf es natürlich Gründe, die zeigen, weshalb Personen, die praktisch überlegen, zumeist zu diesem Urteil gelangen. Eine andere Antwort wäre zu bestreiten, dass solche Fälle nie vorkämen, dass es also durchaus Fälle von irrationalem Lustverzicht gibt. Dies ist nicht ganz so sonderbar, wie es Schälike suggeriert: Zur alltäglichen Erfahrung gehören neben den Genussüchtigen ebenso auch Charaktere, die ein gespaltenes Verhältnis zur eigenen Lust leben, darunter auch solche, die trotz besserer Einsicht auf Lust verzichten. Im Alltag werden diese oft als unfähig beschrieben, richtig zu genießen. Diese Möglichkeit lässt sich auch dann vertreten, wenn man annimmt, dass das Luststreben unter allen Strebearten eine äußerst dominante und zentrale Rolle einnimmt. Schließlich steht dieser Asymmetrie der Willensschwäche auch die gleichartige Asymmetrie der Selbsttäuschung gegenüber: Weshalb täuschen sich Personen tendenziell eher bei unangenehmen Wahrheiten als bei angenehmen? Rascher und hoher Lustgewinn ist ein starkes Motiv, das die willensschwache Person dazu bewegt, die willensschwache Handlung auszuführen entgegen ihrem besseren Urteil. Ebenso wird die Selbsttäuschungskonzeption argumentieren, dass Lustgewinn ein starkes Motiv sei, sich selbst zu täuschen. Weshalb sollte also eine dualistische Konzeption auf diese Art der Erklärung verzichten müssen?

*Viertens*<sup>191</sup> könne eine dualistische Konzeption den Umstand nicht erklären, weshalb Willensschwäche bei Entscheidungen zwischen wirklich wichtigen und relativ unwichtigen Zielen nicht vorkomme. Wir bleiben nicht willensschwach im Bett liegen, wenn dies bedeutete, dass wir damit unsere gesamte Existenz ruinierten. Abgesehen davon, dass solche Fälle zumindest vorstellbar sind, zielt die Frage darauf ab, wie die Diskrepanz zwischen den Handlungsalternativen überhaupt einen Einfluss auf die tatsächliche Entscheidung innerhalb dualistischer Konzeptionen haben kann. Wenn das abschließende beste Urteil und die Intention prinzipiell voneinander

---

<sup>190</sup> Vgl. Schälike 2004, 369f.

<sup>191</sup> Vgl. Schälike 2004, 370f.

verschieden sind, dann heißt dies jedoch nicht, dass die Intention nichts mit dem abschließenden Urteil zu tun haben kann. Die Intention könnte das Urteil integrieren, indem es deren Gründe, es für wahr zu halten, als Gründe übernimmt, die entsprechende Handlung zu intendieren. Die willensschwache Person lässt ihren Willen – ihre Intention, etwas zu tun bzw. ihre Entscheidung – von Gründen bestimmen, die sie evaluativ als minderwertig gegenüber anderen Gründen einstuft. Wirklich wichtige Entscheidungen haben es jedoch an sich, dass sie nicht nur intellektuell, sondern auch affektiv und emotional mit starken Strebungen verbunden sind, Handlungen zu meiden oder sie zu tun. Unter dieser Annahme erscheint es nicht merkwürdig, dass in solchen Situationen Willensschwäche höchst unwahrscheinlich ist, da das bessere Urteil mit starken Strebungen verknüpft ist und so der charakteristische Konflikt der Willensschwäche zwischen besserem Urteil und divergierenden Strebungen gar nicht erst aufkommt.

Für Schälike sind diese vier Umstände dennoch Gründe, die für eine monistische Lösung sprechen, innerhalb derer (vermeintliche) Willensschwäche als Urteilsrevision gepaart mit Mechanismen der Selbsttäuschung zu interpretieren ist. Die willensschwache Person revidiert ihr abschließendes Urteil in der konkreten Situation zugunsten der willensschwachen Alternative und täuscht sich vorher und/oder nachher bezüglich der eigenen praktischen Einstellungen. Die Selbsttäuschung ist notwendig, da die Person ja weiterhin von sich behauptet, entgegen ihrem besseren Urteil gehandelt zu haben. Da innerhalb einer monistischen Konzeption dieses Urteil identisch ist mit der Intention, mit der die willensschwache Person gehandelt hat, und Personen normalerweise unmittelbar wissen, mit welcher Intention sie handeln, müsste die willensschwache Person gleichzeitig zwei einander widersprechende Glaubenszustände bezüglich ihrer praktischen Einstellung haben, was ohne weiteres gemeinhin nicht für möglich gehalten wird. Schälike stellt drei Formen von Selbsttäuschung vor, welche dieses Problem umgehen sollen.<sup>192</sup> Grundlegend für alle drei ist, dass Selbsttäuschung als Wunschdenken verstanden wird, das den Meinungsbildungsprozess irrational beeinflusst.<sup>193</sup>

„Der Selbsttäuscher konfrontiert sich in einem zeitlich ausgedehnten Prozess mit Indizien, die die Ausbildung der erwünschten Meinungen gerechtfertigt erscheinen lassen. Solch ein Prozess der Selbstveränderung ist irrational, insofern es nicht am Interesse an der Wahrheit ausgerichtet ist; er ist aber insofern auch rational, als es dem Subjekt wichtiger erscheint, bestimmte Meinungen zu haben, als dass die Meinungen wahr sind.“ (Schälike 2004, 374)

---

<sup>192</sup> Vgl. auch Schälike 2002, 55ff.; Schälike 2009.

<sup>193</sup> Dieser Prozess der Selbsttäuschung basiert auf Annahmen des *indirekten* doxastischen Voluntarismus: eine Person kann diesem gemäß nur indirekt auf ihren Meinungsbildungsprozess einwirken und nicht direkt. Eine Person kann also nicht einfachhin etwas glauben, nur weil sie es glauben will. Sie muss sich stattdessen selbst mit Evidenzen konfrontieren, welche bewirken, dass sich der gewünschte Glaube einstellt. Der *direkte* doxastische Voluntarismus geht dagegen davon aus, dass Personen ihre Glaubenszustände direkt beeinflussen können. Zur Kritik am direkten doxastischen Voluntarismus vgl. Williams 1970. Vgl. auch die Diskussion um willensschwache Überzeugungen (akrasia of belief) in Kap. I.4.

In der *ersten* Form rationalisiert die willensschwache Person ihre willensschwache Handlung, indem sie diese als berechtigte Ausnahme von einem ansonsten als richtig befundenen Prinzip darstellt.<sup>194</sup> Dafür sucht sie nach Gründen, welche die willensschwache Handlung als eine berechtigte Ausnahme erscheinen lassen. Die willensschwache Person manipuliert demnach ihre praktische Überlegung *vor* der entsprechenden Handlung.

Die *zweite* Form betrifft die Manipulation der Meinungsbildung *nach* vollzogenen willensschwachen Handlungen. Aufgrund des Wunsches nach Aufrechterhaltung eines bestimmten Selbstbildes legt sich die (vermeintlich) willensschwache Person eine falsche, nämlich dualistische Handlungstheorie zurecht, die zwischen Evaluation und Motivation unterscheidet.<sup>195</sup>

„Auf diese Weise entsteht das Alltagskonzept von Willensschwäche, die Vorstellung nämlich, das „wahre“, evaluative Selbst sei bedauerlicherweise nicht „stark“ genug, um die Direktiven der praktischen Vernunft gegen die Dispositionen auf der Ebene des motivationalen Selbst durchzusetzen, auf der die Entscheidungen fallen. Zum Schein ermöglicht dieser Zug eine Teil-Exkulpation, da er es zulässig erscheinen lässt, wenigstens auf der evaluativen Ebene an dem attraktiven Selbstbild festzuhalten.“ (Schälike 2004, 376)

Der Begriff der Willensschwäche sei also selbst ein Produkt von Selbsttäuschungsprozessen: Personen schreiben sich selbst Willensschwäche zu, um sich über ihre wahren Handlungsmotive zu täuschen. Die willensschwache Person täuscht sich selbst, eine Person zu sein, die evaluativ an bestimmten Prinzipien oder allgemein Werturteilen festhalten kann und zugleich diese nicht ins Handeln umsetzt, obwohl sie dies hätte tun können. Ob dies tatsächlich das Alltagskonzept von Willensschwäche ausmacht, dass sich die Person unter Hinweis auf eine Spaltung teilweise selbst entschuldigt, kann jedoch angezweifelt werden: schließlich kommt es vor, dass sich die Person mit ihrer Willensschwäche konfrontiert und sich ihre eigene Willensschwäche vorwirft, anstatt sich selbst mit dem Hinweis auf ihre Willensschwäche zu entschuldigen.

Die *dritte* Form betrifft das bessere Urteil, das die willensschwache Person über ihre Interessenshierarchie fällt. Die Person könnte sich über die Stärke ihres Interesses im Vergleich zu ihren anderen Interessen täuschen:

„Nun können wir zwar das, wofür wir uns interessieren, unmittelbar erkennen, es zeigt sich uns in den evaluativ-motivationalen Bewusstseinszuständen, die mit der Imagination bestimmter Sachverhalte verbunden sind. Da wir uns als endliche Wesen aber nicht unbegrenzt viele Sachverhalte gleichzeitig vorstellen können, wissen wir nicht zu jedem Zeitpunkt alles über unsere Wünsche, sondern nur, ob wir die *situativ imaginierten* Sachverhalte realisiert wünschen.“ (Schälike 2004, 377)

Die willensschwache Person kann demnach aufrichtig meinen, ihre Interessen seien im Einklang mit ihren Urteilen über ihre Prinzipien, und sich dennoch dabei irren, da sie es verabsäumt, sich die entsprechenden Sachverhalte hinreichend gut vorzustellen. Dieses falsche Urteil kann jedoch

---

<sup>194</sup> Vgl. Wolf 1985, 30f.

<sup>195</sup> Vgl. Wolf 1985, 31f.

nur nach oder vor der entsprechend willensschwachen Handlung gefällt werden, da wir in der Situation um unsere Wünsche und Interessen unmittelbar und unfehlbar wüssten.<sup>196</sup>

Aus dieser Beschreibung von Willensschwäche als Selbsttäuschung folgt, dass die willensschwache Person im Moment ihres willensschwachen Handelns kein Bewusstsein über den Konflikt, in dem sie sich befindet, hat. Schließlich *ändert* sie nach Schälke in der Situation ihre Meinung darüber, was am besten zu tun ist, und kann daher aus ihrer Sicht den Konflikt, in dem sie sich befindet, nicht z.B. als Konflikt zwischen gerade aufstrebenden Wünschen und ethischen Prinzipien beschreiben. Es bleibt die Frage bestehen, weshalb das Urteil darüber, was jemand für evaluativ besser hält, ein Urteil über die gegebenen Wünsche oder Interessen der Person sein sollte. Dazu braucht es Argumente, die zeigen, dass evaluative Urteile inhaltlich von Wünschen handeln oder solche Wünsche bloß ausdrücken. Die Diskussion um die Möglichkeit von Willensschwäche geht hier direkt über in Diskussionen über Fragen der Metaethik (Nonkognitivismus und Subjektivismus). Festzuhalten ist, dass Positionen, welche evaluative Urteile als Urteile über tatsächlich vorliegende Wünsche verstehen, oder als Ausdruck solcher Wünsche, synchrone Formen von Willensschwäche konsequenterweise ausschließen müssen. Dies lässt sich anhand des folgenden Argumentes verstehen:

- (P1) Ein situativ bestes Urteil ist ein Urteil darüber, was eine Person am meisten in einer Situation wünscht. (Explikation von ‚situativ bestes Urteil‘)
- (P2) Das, was eine Person am meisten in einer Situation wünscht, ist einer Person unmittelbar epistemisch zugänglich (Epistemische Unmittelbarkeit von Wünschen)
- (K1) Also: Ein situativ bestes Urteil ist ein Urteil über etwas, was einer Person unmittelbar zugänglich ist.
- (P3) Ein Urteil über etwas, was einer Person unmittelbar zugänglich ist, ist wahr. (Gewissheit von Urteilen über epistemisch Unmittelbares)
- (P4) Ein situativ bestes Urteil ist wahr.
- (P5) Eine Person tut das, was sie in einer Situation am meisten wünscht, wenn sie so oder anders handeln kann. (Explikation von ‚etwas am meisten wünschen‘)
- (K2) Also: Eine Person handelt immer entsprechend ihrem situativ besten Urteil, wenn sie so oder anders handeln kann.

---

<sup>196</sup> „Die Quintessenz dieser Überlegungen ist also nicht, dass es Willensschwäche gar nicht gibt, sondern dass sie nicht darin besteht, absichtlich etwas anderes als das zu tun, was man situativ für das Beste hält. Im Modus der Selbsttäuschung hält der Akratiker einige von ihm hochgeschätzte Strebungen für stärker, als sie tatsächlich sind. Sein Wille ist nicht „zu schwach“, um sein Werturteil in die Tat zu setzen, sondern *beide*, Motivation *und* Präferenzen, sind nicht so stark, wie er behauptet. Dieser Befund scheint in einer gewisser Spannung zu meiner These zu stehen, wir hätten ein unmittelbares und unfehlbares Wissen von unseren Wünschen qua intentionalen Einstellungen. Schließt dieses Wissen die Wunschhierarchie nicht mit ein?“ (Schälke 2002, 58)

Dies steht im Widerspruch zu folgender Beschreibung von synchroner Willensschwäche:

- (P1) X fällt ein situativ bestes Urteil.
- (P2) X kann so oder anders handeln.
- (P3) X handelt entgegen dem situativ besten Urteil.

In der Struktur gleicht dieses Argument dem vorigen Argument von Wolf gegen das Verständnis von Willensschwäche als Handeln wider das situativ beste Urteil. Es ist lediglich ergänzt durch eine genauere Analyse dessen, was ein situativ bestes Urteil ist und welchen epistemischen Zugang wir zu diesem haben. Bei Wolf fehlte eine Erklärung dafür, weshalb ein situativ bestes Urteil identisch mit einem Handlungsvorsatz sein soll – außer ihrem Verweis auf die präskriptivistische Interpretation bestimmter Werturteile. Schälike erläutert diesen Zusammenhang mit der Explikation des besten Urteils als Urteil über bzw. Ausdruck von tatsächlich vorliegenden Wünschen einer Person.<sup>197</sup> In dieser ungenauen Formulierung – Urteil über oder *Ausdruck* von – verbirgt sich ein Problem: Ist das evaluative Urteil *expressiv*, also lediglich *Ausdruck* von tatsächlich vorhandenen Wünschen, dann ist unklar, wie sich eine Person selbst über ihre Wünsche in einem evaluativen Urteil *täuschen* könnte, schließlich *handeln* diese Urteile nicht *über* das Vorliegen solcher Wünsche, sondern drücken diese lediglich aus. Ist das evaluative Urteil jedoch *subjektivistisch*, handelt es also über vorliegende Wünsche, dann ist einleuchtend, weshalb eine Person sich selbst täuschen könnte, aber es ist unklar, weshalb ein solches Urteil mit Intentionen, etwas zu tun, gleichzusetzen ist. Hier braucht es eine Theorie, die den Zusammenhang zwischen solchen Urteilen genauer erläutert im Sinne des motivationalen Internalismus. Die Frage, ob die Identifikation von Wünschen mit evaluativen Einstellungen gerechtfertigt ist, ist also eine über die Selbsttäuschungsthese hinausgehende metaethische Frage, bei der Schälike einen handlungstheoretischen und ethischen Internalismus vertritt.<sup>198</sup>

### III.2 Willensschwäche als fehlerhafte Diskontierung

Die Explikation der Willensschwäche als fehlerhafte Diskontierung ist so alt wie die philosophische Diskussion um Willensschwäche selbst. Ich diskutiere diese Form von Theorien anhand von Platons Dialog *Protagoras* sowie George Ainslies moderner Version in *Breakdown of will* (2004). Die Täuschung, der willensschwache Personen unterliegen, betrifft im Unterschied zu Selbsttäuschungstheorien primär nicht eigene Eigenschaften, sondern Eigenschaften der begehrten Gegenstände oder Objekte. Die willensschwache Person irrt sich in ihrer Einschätzung der Eigenschaften der erstrebten Objekte und dies mehr oder weniger systematisch. Diese Theorien

---

<sup>197</sup> Vgl. Schälike 2002.

<sup>198</sup> Vgl. die Diskussion um den motivationalen Internalismus in Kap. III.5.

unterscheiden sich voneinander erstens anhand der Qualitäten, über die sich die willensschwache Person täuscht, und zweitens anhand der Erklärung für die mehr oder weniger systematische Fehlleistung. Als Konsequenz ergeben diese Theorie eine Ablehnung von synchroner, motivationaler Willensschwäche, da die Person im entsprechenden Moment der Willensschwäche ihre Meinung darüber, was am besten zu tun ist, ändert.

### III.2.1 Protagoras: Willensschwäche als fehlende Messkunst

Platon lässt Sokrates in seinem Dialog *Protagoras* gegen die Möglichkeit synchroner Willensschwäche argumentieren. Die genaue Interpretation ist dabei mindestens ebenso umstritten wie die aristotelische Behandlung von Willensschwäche und es ist hier nicht der geeignete Rahmen, um alle möglichen Deutungen durchzugehen und deren Vorzüge und Schwächen gegeneinander abzuwägen.<sup>199</sup> Ich diskutiere hier eine Interpretation, nach der Sokrates die Beschreibung von Willensschwäche als Handeln wider besseres Wissen ablehnt und das entsprechende Handeln als Produkt fehlerhafter Messkunst bzw. Diskontierung ausweist. Der Dialog dreht sich zentral um die Frage der Einheit der Tugenden und der Lehrbarkeit von Tugendhaftigkeit. Sokrates fragt dabei den Sophisten Protagoras, was für ein Wissen er denn genau lehre, wenn er Tugendhaftigkeit gegen teures Geld unterrichte, und ob er meine, dass verschiedene Tugenden bloß verschiedene Namen für ein und dasselbe seien, oder voneinander verschieden.<sup>200</sup> Sokrates argumentiert für die Einheit der Tugenden und die Identifikation von Wissen mit Tugendhaftigkeit. Vor dem für Willensschwäche relevanten Teil des Textes wird die Frage der Trennbarkeit von Tapferkeit und Weisheit aufgeworfen, die Sokrates über den Umweg der Diskussion der Willensschwäche beantwortet. Protagoras vertritt die Ansicht, dass die Tugenden der Weisheit und Tapferkeit getrennt voneinander vorkommen können, d.h. dass eine Person tapfer aber nicht weise oder weise aber nicht tapfer sein könnte. Gegen die erste Möglichkeit wendet Sokrates ein, dass solche Personen tollkühn und nicht tapfer sind, da sie die Gefahr, in der sie sich befinden, nicht wirklich kennen.<sup>201</sup> Weisheit sei demnach eine notwendige Bedingung von Tapferkeit. Die Argumentation gegen Willensschwäche soll nun zeigen, dass Weisheit auch hinreichend ist für Tapferkeit, womit die Äquivalenz von Weisheit und Tapferkeit gezeigt wäre. Im Hintergrund steht die weiterreichende sokratische These, dass Erkenntnis allgemein hinreichend sei für Tugendhaftigkeit und im Umkehrschluss Lasterhaftigkeit auf fehlendem

---

<sup>199</sup> Vgl. u.a. Santas 1966; Vlastos 1969; Segvic 2006; Penner 1997; Gardner 2002; Morris 2006.

<sup>200</sup> Vgl. "Es war aber, wie ich glaube, die Frage diese, ob Weisheit und Besonnenheit und Tapferkeit und Gerechtigkeit und Frömmigkeit, ob dies nur fünf verschiedene Namen für eine Sache sind oder ob jedem dieser Namen auch eine eigene Wesenheit unterliegt und eine eigene Sache, die jede ihre besondere Verrichtung hat, so daß die eine nicht ist wie die andere." (Platon 2009, 349a–b)

<sup>201</sup> Vgl. Platon 2009, 349d–350c.

Wissen beruhe. Ausgangspunkt ist die Ansicht des common sense, dass Personen zuweilen willensschwach handeln:

„Du weißt aber doch, daß die meisten Menschen mir und dir nicht glauben, sondern sie sagen, daß viele, welche das Bessere sehr gut erkennen, es doch nicht tun wollen, obgleich sie könnten, sondern etwas anderes tun. Und so viele ich gefragt habe, was doch die Ursache wäre hiervon, haben mir alle gesagt, von der Lust überwunden oder der Unlust oder irgendeinem unter den Dingen, deren ich vorhin erwähnte [Liebe, Furcht, Zorn], bezwungen, taten die das, die es tun.“ (Platon 2009, 352d–e)

Sokrates argumentiert gegen den common sense auf Basis einer *reductio ad absurdum*.<sup>202</sup> Willensschwäche wird bestimmt als der Fall, dass eine Person durch Lust überwunden wird etwas zu tun, von dem sie weiß, dass es schlechter ist als eine mögliche alternative Handlung. Es folgt eine Argumentation für eine hedonistische Identifikation von ‚ist gut‘ mit ‚ist angenehm‘ bzw. von ‚ist schlecht‘ mit ‚ist unangenehm‘.<sup>203</sup> Ersetzt man diese Begriffe nun in der Beschreibung von angeblicher Willensschwäche, so folgen absurde Konsequenzen. Ersetzt man ‚ist angenehm‘ durch ‚ist gut‘, so kommt man zu dem Schluss, dass eine willensschwache Person etwas Schlechtes tut, weil sie durch das Gute überwunden wird, obwohl sie weiß, dass es schlecht ist. Dies ist aber unmöglich, da wir die Person für ihr willensschwaches Handeln kritisieren, das Gute an der Handlung demnach tatsächlich kleiner sein muss als das Schlechte.

„Daß dies lächerlich ist, wird euch ganz klar werden, sobald wir uns nicht mehr der vielerlei Namen zugleich bedienen wollen, des Angenehmen und Peinlichen und des Guten und Bösen, sondern da sich gezeigt hat, daß dieses nur zweierlei ist, es auch nur mit zwei Worten bezeichnen wollen, zuerst überall durch gut und böse, und dann wieder überall durch angenehm und peinlich. Dieses also festgestellt, sagen wir, daß der Mensch, das Böse erkennend, daß es böse ist, es dennoch tut. Wenn uns nun jemand fragt: Warum denn? So werden wir sagen, weil er überwunden ist. Wovon denn? wird uns jener fragen; wir aber dürfen nicht mehr sagen: Von der Lust, denn die Sache hat nun einen andern Namen bekommen, und statt Lust heißt sie Gutes. Wir antworten also jenem und sagen: Weil er überwunden ist. Wovon denn? fragt er. Von dem Guten, werden wir beim Zeus sagen müssen. Ist nun der, welcher uns fragt, ein Spötter, so wird er lachen und sagen, das ist doch wahrhaftig eine lächerliche Sache, was ihr da sagt, daß ein Mensch das Böse tut, indem er erkennt, daß es böse ist, und obwohl er es nicht tun muß, es dennoch tut, weil er vom Guten überwunden ist! Von einem Guten, wird er fragen, welches wert oder welches nicht wert war, jenes Böse zu überwinden? Offenbar werden wir zur Antwort sagen müssen: Von einem, welches dessen nicht wert war; denn sonst hätte der nicht gefehlt, von dem wir sagen, daß er zu schwach war gegen die Lust.“ (Platon 2009, 355b–d)

Ersetzt man wiederum ‚ist schlecht‘ durch ‚ist unangenehm‘, so gelangt man zu dem Schluss, dass die willensschwache Person etwas tut, weil es angenehm ist, obwohl sie weiß, dass es unangenehm ist für sie. Dies ist wiederum absurd, da auf Basis eines psychologischen Hedonismus

---

<sup>202</sup> Vgl. Brickhouse & Smith 2006, 266–269.

<sup>203</sup> Vgl. Platon 2009, 353c–354e. Es ist umstritten, ob Sokrates diese Identifikation nur *ad hominem* annimmt, d.h. nur als eine der Voraussetzungen seines Gegenübers für die Dauer des Gesprächs akzeptiert, oder ob er sie wirklich vertritt. Vgl. Müller 2009, Kap. II.2.1; Brickhouse & Smith 2006, 267.

eine Person von zwei Handlungen immer diejenige wählt, die lustvoller bzw. angenehmer ist. Das gesamte Argument kann folgendermaßen verstanden werden:<sup>204</sup>

- (P1) Manchmal tut eine Person P eine Handlung F, weil F-en angenehm ist für P, obwohl P weiß, dass F-en schlecht ist für P. (Definition von Willensschwäche)
- (P2) Wenn etwas gut (schlecht) ist, dann ist es angenehm (unangenehm), und wenn etwas angenehm (unangenehm) ist, dann ist es gut (schlecht). (hedonistische Identifikation)
- (K1) Also: Manchmal tut eine Person P eine Handlung F, weil F-en gut ist für P, obwohl P weiß, dass F-en schlecht ist für P. (Ersetzung von ‚ist angenehm‘ durch ‚ist gut‘)
- (K2) Aber: Das Gute an F-en kann das Schlechte an F-en nicht überwiegen wenn die Person einen Fehler begeht indem sie F-t.
- (K3) Also: Manchmal tut eine Person P eine Handlung F, weil F-en angenehm ist für P, obwohl P weiß, dass F-en unangenehm ist für P. (Ersetzung von ‚ist schlecht‘ durch ‚ist unangenehm‘)
- (K4) Aber: Personen wählen immer die Handlung, von der sie glauben, dass sie am angenehmsten ist (psychologischer Hedonismus)

Der dem Argument zugrunde liegende psychologische Hedonismus weist jedoch nach Sokrates auch auf die eigentliche Lösung für das Rätsel hin, weshalb Personen willensschwach handeln: die Personen können die zeitlich nahen und zeitlich fernen Güter und Übel nicht korrekt gegeneinander abwägen. Ebenso wie räumlich ferne Gegenstände klein und nahe groß erscheinen, so erscheinen zeitlich ferne Gegenstände weniger attraktiv und zeitlich nahe attraktiver, als es ihrer wahren Größe entspricht.<sup>205</sup> Willensschwachen Personen fehlt es daher an einer Messkunst, Lust und Unlust richtig abzuwägen, und da diese Messkunst eine Form von Wissen oder Weisheit darstellt, mangelt es der willensschwachen Person an Weisheit – sie handelt damit nicht wider ihr besseres Wissen, sondern aus fehlender Weisheit. Neben den problematischen Voraussetzungen des Argumentes, dem psychologischen Hedonismus und die hedonistische Identifikation von Wertprädikaten mit Prädikaten über Lust- bzw. Unlustzuständen, bleibt vor allem unklar, wie innerhalb eines solchen Modells motivationale Konflikte einer Person überhaupt modelliert werden können. Gemäß der Theorie sind Evaluation und Motivation stets im Einklang zueinander, was echte Konflikte zwischen verschiedenen Strebungen grundsätzlich ausschließt.<sup>206</sup> Das Modell, dass geeignet ist, um dia-

---

<sup>204</sup> Vgl. Müller 2009, 68.

<sup>205</sup> Vgl. Platon 2009, 356b–d.

<sup>206</sup> „Welche Art von innerem Konflikt möglich ist, hängt [...] nicht unwesentlich von den psychologischen Grundannahmen ab, unter denen die Diskussion geführt wird. Dies zeigt schon die sokratische Erörterung im *Protagoras*: Nimmt man an, dass die Seele eine in sich ungeteilte Entität ist, die durch rationale Überlegung ein im

chrone Willensschwäche zu erläutern, stößt dadurch bei der Erläuterung und Konzeption synchroner Willensschwäche und anderer synchron stattfindender Konflikte an seine Grenze.

### III.2.2 George Ainslie: Willensschwäche als hyperbolische Diskontierung

Innerhalb der Ökonomie und der Theorie intertemporaler Entscheidungen kann die platonische Messkunst als Kunst der Diskontierung (zukünftigen) Nutzens interpretiert werden. Eine intertemporale Entscheidung ist eine Entscheidung einer Person über verschiedene Optionen zu einem Zeitpunkt  $t$ , die erst zu davon verschiedenen Zeitpunkten  $t + d$  wirksam werden. Innerhalb der Nutzentheorie geht es dabei also um die Abwägung bzw. den Abtausch (trade-off) von erwartetem Nutzen, der zu verschiedenen Zeitpunkten auftritt, bzw. zwischen gegenwärtigem und zukünftigem Nutzen. Dies sind Entscheidungen wie z.B. die, ob ich lieber heute ins Kino gehe oder doch für die bald anstehende Prüfung lerne, oder ob ich für eine große Reise im nächsten Sommer spare oder doch lieber bereits nächste Woche Skifahren gehe.

Diskontierung wird in diesem Zusammenhang verstanden als Berechnung des Nutzens eines Gutes zu einem Zeitpunkt  $t + d$  (also Wert  $V(u, t + d)$ ) aus der Perspektive eines davon verschiedenen Zeitpunktes  $t$  (also  $V(u, t)$ ).<sup>207</sup> Trifft eine Person eine intertemporale Entscheidung, so wägt sie den mit den verschiedenen Optionen verbundenen Nutzen gegeneinander ab, wozu sie den Nutzen relativ zum seinem Zeitpunkt diskontiert. Stehe ich z.B. vor der Entscheidung, ob ich lieber morgen 100 € bekomme oder erst in einer Woche 120 €, dann wäge ich den Nutzen der morgigen 100 € gegen den Nutzen von 120 € in einer Woche ab.<sup>208</sup> Das Resultat der Abwägung wird unter anderem davon abhängen, wie viel ich dem Umstand beimesse, dass der Nutzen der beiden Optionen zu verschiedenen Zeiten verfügbar ist. Ziehe ich die morgigen 100 € den erst in einer Woche verfügbaren € 120 vor, dann *kann* dies bedeuten, dass mir eine Woche Wartezeit mehr als 20 € wert ist.<sup>209</sup>

---

Handeln grundsätzlich umgesetztes Urteil über das in der jeweiligen Situation Beste hervorbringt, sind alle stärkeren Formen des inneren Konflikts ausgeschlossen. Urteil und Begierde, Evaluation und Motivation können nicht synchron auseinanderklaffen, sondern lediglich unter dem Einfluss äußerer bzw. körperlicher Aspekte [...] schwanken bzw. oszillieren. Das Urteil kann sich temporär ändern, aber ein simultaner Konflikt ist ausgeschlossen, so dass in diesem Modell nur eine diachrone Form von Willensschwäche akkomodiert werden kann.“ (Müller 2009, 201)

<sup>207</sup> Vgl. Loewenstein & Prelec 1992; Lind 2010; Rachlin 2000, 30ff. „A discount rate ( $\rho$ ) can be defined as the extra fraction of a pay-off needed to make an agent indifferent between satisfying some desire now and satisfying it in one period.“ (Heath 2003, 256) „A time discount function is a mathematical expression of the reduction in value of some commodity due to its delay.“ (Rachlin 2000, 30)

<sup>208</sup> Ich verwende aus Gründen der besseren Anschaulichkeit in den Beispielen Geldwerte statt Nutzeinheiten.

<sup>209</sup> ‚Kann‘ deshalb, da vielleicht andere Faktoren es besonders dringlich machen, dass ich die € 100 schon morgen bekomme z.B. wenn ich dringend Schulden begleichen muss oder ein Medikament sofort benötige.

Es lassen sich verschiedene Gründe unterscheiden, weshalb eine Person zukünftigen bzw. verzögert eintretenden Nutzen geringer oder höher bewertet als gegenwärtigen Nutzen:<sup>210</sup> a) Risiko bzw. Unsicherheit bezüglich des Eintretens von verschiedenen Optionen: die subjektive oder objektive Wahrscheinlichkeit des Akteurs über mögliche Weltverläufe; b) die vorhersehbare Veränderung von Präferenzen über die Zeit: die Einschätzung des Akteurs über Änderungen in seiner Präferenzstruktur über die Zeit hinweg; c) Inflation: die sinkende Kaufkraft von Geld; d) abnehmender Grenznutzen: der sinkende Grenznutzen bzw. zusätzliche Nutzen eines Gutes mit seinem fortschreitenden Konsum; e) Opportunitätskosten: Kosten für den Konsumverzicht z.B. aus Sicht des Kreditgebers; f) Nutzen aus Antizipationen: Nutzen aus der Vorstellung des Eintretens eines Ereignisses (Warten auf das Christkind oder den Zahnarzt); g) Gewohnheiten: Änderungen von Konsumgewohnheiten.<sup>211</sup> Von diesen Faktoren sollte die sogenannte *pure Zeitpräferenz* unterschieden werden, also die reine Bevorzugung eines Nutzens bezüglich des Zeitpunktes und der Verzögerung seines Eintretens. Es ist theoretisch und methodisch schwierig die vorigen Faktoren von *reiner Zeitpräferenz* zu trennen und empirisch zu bestimmen. Außerdem können diese Faktoren auf unterschiedliche Weise in die jeweiligen Modelle integriert werden. Das Risiko, das mit bestimmten Ereignissen verbunden ist, kann z.B. direkt in die Berechnung des Erwartungsnutzens einbezogen werden, wobei noch die Risikofreudigkeit bzw. Aversion gegenüber Risiko als zusätzlicher Faktor miteinbezogen werden kann. Zusätzlich müssen deskriptive von normativen Überlegungen bezüglich des ‚richtigen‘ Diskontierens auseinander gehalten werden: die Frage, wie Personen (oder andere Lebewesen) *tatsächlich* zeitliche Verzögerung bewerten ist eine andere als die Frage, wie sie es rationalerweise tun sollten.

Neben den Gründen für die Diskontierung unterscheidet man mögliche Gegenstände der Diskontierung: Güter, Nutzen und Wohlbefinden sind die am häufigsten genannten. Ob es zulässig ist, Wohlbefinden selbst zu diskontieren, ist eine strittige Frage, obwohl die Gründe eher dafürsprechen, dies nicht zu tun: Das gleiche Ausmaß an Wohlbefinden sollte zu jedem Zeitpunkt gleich viel wert sein.<sup>212</sup> Wenn Nutzen als Wohlbefinden verstanden wird, so trifft dieses Argument gleichfalls auf den Nutzen zu. Wird Nutzen jedoch verstanden als die Fähigkeit eines bestimmten Guts zu einer bestimmten Zeit, Wohlbefinden zu spenden, dann könnten ver-

---

<sup>210</sup> Vgl. Frederick & Loewenstein & O'Donoghue 2002.

<sup>211</sup> Anhand des Beispiels: Ich könnte mir a) nicht sicher sein, die 120 € in einer Woche zu bekommen; b) mir könnte in einer Woche allgemein Geld weniger wichtig sein; c) die 120 € könnten in einer Woche weniger Kaufkraft haben; d) 120 € könnten in einer Woche weniger Nutzen stiften, wenn ich weiß, dass ich inzwischen ein großes Erbe antrete; e) ich könnte die 100 € gewinnbringend investieren und hätte so in einer Woche mehr als 120 €; f) die Erwartung, Geld zu bekommen, könnte ich selbst als angenehm empfinden (oder als quälend), weshalb ich es vorziehe, das Geld später (bzw. früher) zu erhalten; g) ich könnte mich an ein bestimmtes Konsumlevel gewöhnen, sodass mir die 120 € nicht mehr so viel ausmachen wie heute.

<sup>212</sup> Vgl. Broome 1994, 147–151.

schiedene Gründe dafürsprechen, Güter dieser Art bezüglich des Zeitpunktes zu diskontieren, an dem sie konsumiert werden können. Unter der Annahme einer wachsenden Wirtschaft mit gleichzeitiger Erhöhung der Produktivität werden bestimmte Güter in Zukunft billiger produziert werden, weshalb sie aus heutiger Sicht weniger wert sind. Oder anders ausgedrückt mittels der Theorie des abnehmenden Grenznutzens: Da ich mir für denselben Betrag in Zukunft mehr von einem bestimmten Gut kaufen kann, der Grenznutzen dieser Güter jedoch abnehmend ist, sind diese Güter aus heutiger Perspektive weniger wert als gleichartige gegenwärtige Güter. Dies ist nicht zu verwechseln mit der These, den Nutzen selbst, den ich in Zukunft durch den Konsum dieser Güter erlange, abzuwerten: Die Güter spenden einfach weniger Nutzen in Zukunft als jetzt.

Ein klassisches Modell für die Diskontierung entwickelte Paul Samuelson (1937). In diesem Modell werden die verschiedenen Faktoren für die Diskontierung in einen Faktor, den Diskontierungsfaktor, zusammengefasst. Im Grunde handelt es sich um eine einfache und elegante mathematische Modellierung der Diskontierung von Konsumbündeln von Gütern über die Zeit hinweg.<sup>213</sup>

$$U^t(c_t, \dots, c_T) = \sum_{k=0}^{T-t} D(k)u(c_{t+k})$$

$$\text{where } D(k) = \left( \frac{1}{1 + \rho} \right)^k$$

(Frederick & Loewenstein & O'Donoghue 2002, 355)

Angenommen, ich diskontiere täglich mit  $\rho = 0,1$ , also 10 %, so wären die 100 €, die ich morgen erhalte, aus heutiger Sicht  $100 * 1/(1+0,1)^1 = 90,90$  €, die 120 € in einer Woche nach diesem Modell lediglich  $120 * 1/(1+0,1)^6 = 67,73$  € wert.<sup>214</sup> Entscheidend an dieser Art der Diskontierung ist, dass sie exponentiell erfolgt und keine dynamischen Inkonsistenzen zulässt: wenn ich zu einem Zeitpunkt eine Option einer anderen vorziehe, so tue ich dies zu allen Zeitpunkten. Wenn ich es heute vorziehe, erst in einer Woche 120 € zu erhalten anstatt bereits morgen 100 €, dann ziehe ich dies auch morgen vor und ändere nicht meine Meinung, nur weil sich die Wartezeit um einen Tag verkürzt hat. Dabei gebe ich zu bedenken, dass das Verhältnis dieser Formeln zur tatsächlichen Diskontierungsleistung von Individuen zumindest interpretationsbedürftig ist. Dies

<sup>213</sup> Also: Der Nutzen (U) zum Zeitpunkt t eines Güterbündels ( $c_t, \dots$ ) über den Zeitraum (t bis T) ist gleich der Summe der einzelnen Nutzeinheiten zu den Zeitpunkten  $u(c_{t+k})$  multipliziert mit dem Diskontierungsfaktor  $D(k)$ . Ainslie gibt eine einfachere exponentielle Formel an, die nur ein einzelnes Gut diskontiert: „Value = Value at no delay x (1 - Discount rate)<sup>Delay</sup>.“ (Ainslie 2005b, 636)

<sup>214</sup>  $U^0(c_0, c_1) = D(0) * u(c_0) + D(1) * u(c_1) = 0 + (1/(1 + 0,1))^1 * 100$ .  $U^0(c_0, c_1, c_2, c_3, c_4, c_5, c_6) = D(0) * u(c_0) + D(1) * u(c_1) + \dots + D(6) * u(c_6) = 0 + 0 + 0 + 0 + 0 + 0 + 0 + (1/(1 + 0,1))^6 * 120$ .

unter anderem deshalb, da sonst der Vorwurf einer unnötigen Intellektualisierung von menschlichen Aktivitäten im Raum steht. Vordergründig *beschreiben* diese und andere Formeln lediglich Prozesse der Wertschätzung von Gütern bezüglich ihres zeitlichen Auftretens.<sup>215</sup> Die Individuen kennen dabei normalerweise diese Formeln nicht und verwenden sie dementsprechend auch nicht, um zu diskontieren. Andererseits werden ökonomisch geschulte Personen aktiv und explizit verschiedene Diskontierungsmethoden und Diskontierungsraten ihren ökonomischen Tätigkeiten zugrunde legen. Hier entwickeln Theorien des Diskontierens natürlicherweise eine präskriptive, also handlungsanleitende und handlungsvorschreibende Funktion. Die Aufgabe, rein deskriptiv zu beschreiben und mathematisch zu modellieren, wie Personen tatsächlich diskontieren, hat demnach zwei Gegenstände: auf der einen Seite die implizit tatsächliche Weise des Diskontierens, auf der anderen die ökonomisch explizite Weise mit Hinblick auf Standards des Diskontierens. Von dieser zweifach deskriptiven Betrachtung wiederum verschieden ist die normative Frage, wie denn rationalerweise richtig diskontiert werden *sollte*.

In welchem Zusammenhang stehen nun diese Überlegungen über das Konzept der Diskontierung zum Phänomen der Willensschwäche? Bei Platon spielt der Gedanke eine Rolle, dass willensschwache Personen den Wert zukünftiger Güter nicht richtig einschätzen. Es fehlt ihnen an der korrekten Messkunst. Diese Messkunst ist in moderner Terminologie die Kunst der Diskontierung, und für George Ainslie ist Willensschwäche die Folge bzw. Ausdruck einer bestimmten Art von Diskontierung.

„*Akrasia* ist nichts anderes als Maximierung der erwarteten und mit Hilfe einer stark gebogenen Kurve diskontierten Belohnung.“ (Ainslie 2005a, 154)

Seine These ist dabei, dass Personen (und andere Lebewesen) nicht exponentiell, sondern hyperbolisch diskontieren.<sup>216</sup> Eine einfache mathematische Modellierung dieser Art der Diskontierung ist folgende:<sup>217</sup>

$$D(t) = 1/(1 + t)$$

Im vorigen Beispiel bedeutet dies, dass die morgigen 100 € aus heutiger Sicht  $100 \text{ €} / (1 + 1) = 50 \text{ €}$ , die 120 € in einer Woche aus heutiger Sicht  $120 \text{ €} / (1 + 6) = 17,14 \text{ €}$  wert sind. Um zusätzlich Differenzen in der Art der Diskontierung zwischen verschiedenen Individuen

<sup>215</sup> Vgl. die Kritik von Howard Rachlin an Ainslie: „But hyperbolic discount functions are most usefully conceived not as internal forces prior to behavior, efficiently causing behavior (that may or may not be inhibited), but as descriptions or summaries of actual overt choices. If anyone has hyperbolic discount functions it is the scientific observer, not the actor.“ (Rachlin 2005, 659)

<sup>216</sup> „Robust evidence has indicated that the basic function by which all vertebrates devalue delayed events is hyperbolic.“ (Ainslie 2005b, 649)

<sup>217</sup>  $D(t)$  ist der Diskontierungsfaktor,  $t$  die zeitliche Verzögerung. „Value = Objective value / (1 + Delay)“ (Ainslie 2004, 207)

abzubilden, kann diese Formel um einen Faktor ( $\alpha$ ) erweitert werden:  $D(t) = 1/(1 + \alpha * t)$ .<sup>218</sup> Ainslie veranschaulicht den Unterschied der beiden Diskontierungsarten anhand ihrer entsprechenden mathematischen Kurven:

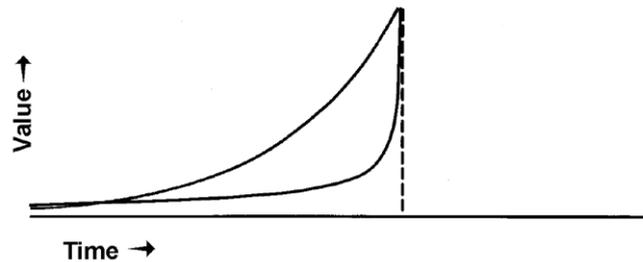


Figure 1. An exponential discount curve and a hyperbolic (more bowed) curve from the same reward. As time passes (rightward along the horizontal axis), the motivational impact – the *value* – of her goals gets closer to its undiscounted size, which is depicted by the vertical line.

(Ainslie 2004, 31)

Die hyperbolische Kurve ist stärker gekrümmt als die exponentielle, sie fällt also anfangs stärker und in weiterer Folge geringer ab. Dies bedeutet, dass eine hyperbolische Diskontierung kurze Verzögerungen gegenüber langen unverhältnismäßig stärker berücksichtigt.<sup>219</sup> Zentral für Willensschwäche ist, dass die hyperbolische im Unterschied zur exponentiellen Diskontierung die Beschreibung dynamischer Inkonsistenzen zulässt. Wenn ich heute die 120 € in einer Woche bevorzuge, doch am morgigen Tag doch die 100 € annehme, *obwohl alle anderen Faktoren gleich geblieben sind*, so zeige ich dynamische Inkonsistenz, da ich dieselbe Entscheidung zu verschiedenen Zeitpunkten unterschiedlich treffe. Analog verlaufen Fälle, die augenscheinlich mehr mit Willensschwäche zu tun haben: die heute feierlich entschiedene Diät wird bereits am nächsten Tag gebrochen, ebenso der Vorsatz, heute die allerletzte Zigarette zu rauchen. Das klassische Standardmodell kann diese Fälle noch beschreiben, indem sie einfach verschiedene Diskontierungsraten für verschiedene Güter annimmt:

„Der einleuchtendste Vorschlag besagt, dass die Menschen bei der Diskontierung verschiedener Arten von Belohnungen verschiedene Sätze zugrunde legen. Wenn ich das Trinken mit 40% pro Tag diskontiere, aber die Vermeidung eines Katers am folgenden Tag (Wert 120 Nutzeinheiten) mit 20%, betrüge der Nettonutzen einer sofortigen Zecherei  $100 - (120 \times 80\%)$  bzw. 4 Nutzeinheiten. Aber einen Tag vorher beliefe sich der Nettonutzen der Zecherei auf  $(100 \times 60\%) - (120 \times 80\% \times 80\%)$ , also  $60 - 76,8$  oder -16,8 Nutzeinheiten. Demnach würde ich trinken, wenn die Möglichkeit dazu unmittelbar gegeben wäre, aber nicht im Fall einer Verzögerung.“ (Ainslie 2005a, 142)

<sup>218</sup> Vgl. Frederick & Loewenstein & O'Donoghue 2002, 360. Dieser Faktor ist vergleichbar mit der Diskontierungsrate  $\rho$  der exponentiellen Diskontierung.

<sup>219</sup> „In contrast to exponential curves, hyperbolic discount curves depict a strong but temporary tendency to prefer smaller and sooner (SS) rewards to larger and later (LL) ones, in the period just before an SS reward is due.“ (Ainslie 2005b, 637)

Das Problem ist, dass dynamische Inkonsistenzen auch in Fällen auftreten, bei denen die Art der Belohnung dieselbe ist, wie im Falle von Geld, das zu verschiedenen Zeitpunkten zur Verfügung steht, und somit die Hypothese verschiedener Diskontierungsraten für verschiedene Güter nicht greift.

„Wenn ich einen ganzen Raum voller Leute bitte, sie sollten sich vorstellen, sie hätten bei einem Wettbewerb gewonnen und könnten zwischen einem sofort einlösbaren, gedeckten Scheck über 100 Euro und einem auf später datierten und erst nach drei Jahren einlösbaren Scheck über 200 Euro wählen, antwortet normalerweise mehr als die Hälfte der Personen, sie hätten lieber gleich 100 Euro. Wenn ich dann fortfahre und frage, wie es mit 100 Euro nach sechs Jahren und 200 Euro nach neun Jahren steht, entscheiden sich fast alle für die 200 Euro. Dabei handelt es sich jedoch um dieselbe Wahl, die aus einem sechs Jahre längeren Abstand gesehen wird.“ (Ainslie 2005a, 146)

Nach exponentieller Diskontierung sind solcherart Inkonsistenzen ausgeschlossen. Hyperbolische Diskontierung eignet sich demnach zur Beschreibung von Fällen dynamischer Inkonsistenz, vor allem von Fällen, in denen kleinere, aber früher verfügbare Belohnungen größeren, dafür erst später verfügbaren vorgezogen werden. Folgende Tabelle ist eine Gegenüberstellung der exponentiellen und hyperbolischen Diskontierung.

	Heute	3 Jahre	6 Jahre	9 Jahre
exponentiell ( $\rho = 0,4$ )	100 € / <b>100 €</b>	200 € / 72,88 €	100 € / <b>13,28 €</b>	200 € / 9,68 €
hyperbolisch ( $\alpha = 0,4$ )	100 € / <b>100 €</b>	200 € / 90,90 €	100 € / 29,41 €	200 € / <b>43,47 €</b>

Links neben dem Trennstrich ‚/‘ sind die tatsächlich zu diesem Zeitpunkt in Aussicht gestellten Belohnungen eingetragen, rechts dieselben diskontierten Beträge. Während die exponentiell diskontierende Person immer die 100 € wählt, wählt die hyperbolisch diskontierende Person die 100 € bar auf die Hand, doch die 200 € in 9 Jahren.

Denselben Sachverhalt veranschaulicht Ainslie folgendermaßen:

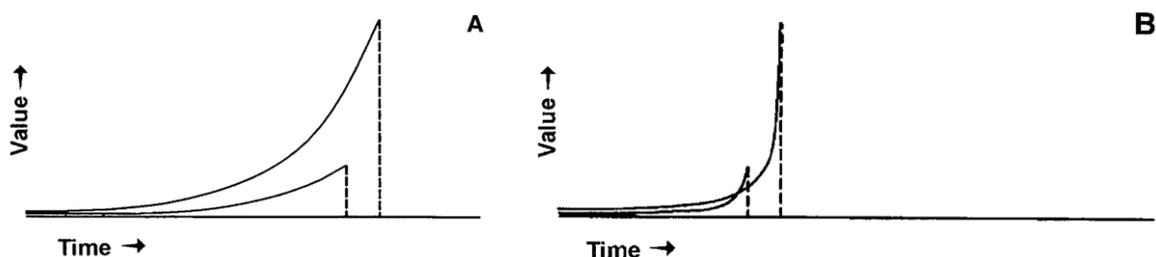


Figure 1.A. Conventional (exponential) discount curves from a smaller-sooner (SS) and a larger-later (LL) reward. At every point their heights stay proportional to their values at the time that the SS reward is due. B. Hyperbolic discount curves from an SS and an LL reward. The smaller reward is temporarily preferred for a period just before it's available, as shown by the portion of its curve that projects above that from the later, larger reward.

(Ainslie 2005b, 636)

Mathematisch lässt sich die exponentielle Formel durch Hinzufügung von Ausnahmen eventuell anpassen, um mit diesen Fällen zurecht zu kommen.<sup>220</sup> Entscheidend ist jedoch die Frage, ob man dynamische Inkonsistenzen als deskriptiven Befund menschlichen (und tierischen Verhaltens) akzeptiert oder nicht.<sup>221</sup> Philosophisch gesehen ist das letztlich eine empirische Frage, die durch bessere Experimente und Versuchsanordnungen zu entscheiden ist. Dagegen ist die Frage philosophisch interessant, welche Art zu diskontieren rational ist. Dies kann nur beantwortet werden, wenn klar ist, welche Art von Rationalität hier gemeint ist. Ein pragmatisches Argument für die Irrationalität hyperbolischen Diskontierens ist, dass eine solch diskontierende Person von einer anderen, exponentiell diskontierenden Person, als Geldpumpe (money pump)<sup>222</sup> missbraucht werden könnte: die hyperbolisch diskontierende Person verliert ständig Ressourcen, ohne ihre eigene Position gemessen an ihren eigenen Präferenzen wirklich zu verbessern.

„Frau Exponential könnte beispielsweise den Wintermantel von Frau Hyperbel in jedem Frühling billig erwerben, weil der Abstand bis zum nächsten Winter die Wertung von Frau H. stärker biegen würde als die Wertung von Frau E. Frau E. könnte dann den Mantel in jedem Herbst wieder an Frau H. verkaufen, sobald das Herannahen des Winters die Mantel-Wertung von Frau H. in die Höhe treibt. Aufgrund dieses mathematischen Musters wird nur eine exponentielle Diskontkurve jemanden vor der Ausbeutung durch eine andere Person schützen, die sich einer Exponentialkurve bedient.“ (Ainslie 2005a, 144)

Unter dem Aspekt nutzenmaximierender, instrumenteller Rationalität verhalten sich hyperbolisch diskontierende Personen irrational, da sie langfristig gesehen ihre Präferenzen nicht optimal erfüllen. Die disproportionale Bevorzugung geringerer gegenüber längerer Wartezeit lässt sie immer wieder gegen ihre eigenen, langfristigen Pläne zugunsten kurzfristiger Belohnungen handeln und damit über die längere Zeit betrachtet ein suboptimales Ergebnis erzielen.<sup>223</sup> Grundlegend für diese normative Bewertung ist jedoch die Annahme eines gleichbleibenden Subjekts dieser Handlungen: die Idee einer gleichbleibenden Person, die Nutznießer der

---

<sup>220</sup> Für quasi-exponentielle Formeln vgl. Frederick & Loewenstein & O'Donoghue 2002, 366ff; Green & Myerson 1993.

<sup>221</sup> Verschiedene Studien zu tierischem Verhalten legen diesen Befund nahe, vgl. Rachlin 2000, 36ff. So zeigen z.B. Tauben in einschlägigen Experimenten ebenso ein Verhalten, das als Präferenzwandel aufgrund sich kreuzender Diskontierungsraten verstanden werden kann. Zu einer Zeit  $t_0$  bevorzugen die untersuchten Tauben 2 Nahrungseinheiten mit einer Verzögerung von 14 Sekunden zu erhalten statt nur einer mit einer Verzögerung von 10 Sekunden. Nach Ablauf von 10 Sekunden ( $t_1$ ) wird den Tauben die Möglichkeit eingeräumt, ihre ‚Entscheidung‘ zu ändern und doch sofort nur eine Nahrungseinheit zu erhalten statt weitere 4 Sekunden auf die ursprünglich festgelegten 2 Nahrungseinheiten zu warten. Die meisten Tauben ändern daraufhin ihre Präferenz und wählen die sofortige Belohnung. Die empirische Basis und Aussagekraft verschiedener Experimente zur Diskontierung ist jedoch keineswegs unumstritten, auch wenn Ainslie sehr selbstbewusst von eindeutigen Ergebnissen und Befunden schreibt. Vgl. Harrison & Lau 2005, 657ff.

<sup>222</sup> Dies ist eine Variante eines Money-Pump-Argumentes. In seiner klassischen Version wird es vorgestellt von Ramsey 1926.

<sup>223</sup> Ainslie untersucht auch mögliche Kehrseiten von Willensstärke, also von exponentieller, an langfristigen Interessen orientierter Diskontierung, vgl. Ainslie 2004, Part III. Näher gehe ich auf die Frage nach der Rationalität bzw. Irrationalität von Willensschwäche in Kap. VI.1 ein.

Ergebnisse ihrer kurz- und langfristigen Pläne ist. Für unser alltägliches Vorverständnis von uns als Personen sind sowohl das Extrem einer allzeit gleichbleibenden Person, wie das gegenteilige Extrem einer sich ständig wandelnden ‚Person‘ – ob hier überhaupt noch von Person gesprochen werden kann, ist fraglich – fremd. Als lernende Wesen sind Personen primär darauf ausgelegt, sich selbst zu wandeln und damit einhergehend auch ihre jeweiligen Interessen und Präferenzen anzupassen und zu modifizieren. Die Frage nach der Rationalität von Präferenzwandel kann somit letztlich nur auf der Basis einer Vorstellung davon beantwortet werden, was es heißt, eine Person zu sein. Bei der hyperbolischen Form der Diskontierung ist jedoch zu beachten, dass der kurzfristige Präferenzwandel einzig von der zeitlichen Entfernung der Person von dem jeweiligen Gegenstand abhängt, also pure Zeitpräferenz betrifft, und alle anderen Faktoren, die unbestritten gute Gründe sein können, Präferenzen zu ändern, idealiter ausgeschlossen sind.

Aus diesen Überlegungen ergibt sich folgendes Problem: Auch wenn die These der hyperbolischen Diskontierung mit dem Phänomen der Willensschwäche und anderen Formen dynamischer Inkonsistenzen gut zurechtkommt, droht die Gefahr, dass plötzlich rationales Verhalten unerklärlich wird. Weshalb handeln Personen, wenn sie eigentlich hyperbolisch diskontieren, zuweilen entgegen dieser natürlichen Tendenz und verhalten sich gemäß dem rationalen, exponentiellen Schema? Die Herausforderung der Willensschwäche wird dementsprechend geradewegs umgedreht: Bestand diese in der Erklärung und Beschreibung irrationalen Verhaltens, ist jetzt die Erklärung und Beschreibung rationalen Verhaltens problematisch.<sup>224</sup>

„Das hyperbolische Diskontieren verlagert das Hauptproblem, vor dem die Nutzentheorie steht. Die Erklärung kurzfristiger Entscheidungen bringt uns nicht mehr in Verlegenheit. Jetzt müssen wir erklären, wie die Menschen jene adaptiven Steuerungsmöglichkeiten lernen, mit deren Hilfe sie sich wie Bankangestellte verhalten.“ (Ainslie 2005a, 154)

Diese Erklärung verlangt dabei eine grundlegende Änderung des traditionellen Personenbegriffs. Die Person ist innerhalb dieses neuen Modells selbst bloß ein Bündel verschiedener kurzfristiger und langfristiger Interessen, die gegeneinander um ihre jeweilige maximale Erfüllung ringen. Aus der Vorstellung eines einheitlichen Selbst, das willensschwach agiert, wird die Vorstellung eines Marktplatzes an Interessen, auf dem die Interessen strategisch um ihre Durchsetzung bemüht sind.

„Ein Akteur, der Belohnungen hyperbolisch diskontiert, ist kein geradliniger Wert-Taxierer, wie man es bei einem exponentiellen Diskontierer unterstellt. Vielmehr besteht er aus einer Reihe von Taxierern, deren

---

<sup>224</sup> „For Ainslie locates akrasia in reversals of preference that occur whenever an agent comes close to a tempting, lesser reward. Ainslie’s explanation of this phenomenon is very original in that it is based on the idea that broad akrasia is the rule, whereas its opposite – enkrateia or strength of will – is the exception.“ (Peijnenburg 2005, 658) Diese Umkehrung ist typisch für die Diskussion um Willensschwäche: je nach Perspektive ist das jeweils konträre Phänomen klärungsbedürftig.

Schlussfolgerungen auseinandergelassen. Mit verstreichender Zeit verändert sich ihr gegenseitiges Verhältnis und wechselt von Kooperation im Hinblick auf ein gemeinsames Ziel zur Konkurrenz um einander wechselseitig ausschließende Ziele. Wenn Odysseus im Hinblick auf die Sirenen Pläne schmiedet, muss er den den Sirenen lauschenden Odysseus als eine getrennte Person behandeln, die nach Möglichkeit beeinflusst werden muss und deren Handeln man, wenn keine Beeinflussung möglich ist, vorbeugen muss.“  
(Ainslie 2005a, 155)

Die Koordination und Integration verschiedener und einander widersprechender Interessen wird darüber hinausgehend nur durch die Begrenztheit der verfügbaren Ressourcen geleistet.<sup>225</sup> Ainslie sieht in der Person das Gefäß für eine Population von Interessen, die strategisch und nutzenmaximierend interagieren, und dementsprechend schlägt er die Erweiterung und Verfeinerung der ökonomischen Analyse auf eine interpersonale Ebene vor (Piko-Ökonomie), auf der die betreffende Art von Transaktionen – der Abtausch und Ausgleich von Interessenserfüllungen – vonstattengehen.

Ich fasse die bisherigen Überlegungen kurz zusammen: Auf der Basis von empirischen Forschungen stellt Ainslie fest, dass Personen (und auch andere Lebewesen) die Aussicht auf zukünftige Belohnungen nicht gemäß einer Standardtheorie rationalen Verhaltens, der exponentiellen Methode, sondern gemäß einer alternativen Methode, der hyperbolischen, bewerten bzw. diskontieren. Dies ermöglicht ihm, bisher in die allgemeine Nutzentheorie nur schlecht bzw. nicht integrierende Phänomene irrationalen Verhaltens wie Willensschwäche, Sucht und allgemein Formen dynamischer Inkonsistenz zu beschreiben und zu erklären. Für Willensschwäche bedingt dies eine Umkehrung der Problemstellung, da nun rationales, an langfristigen Interessen orientiertes Handeln von Personen unverständlich zu werden droht. Schließlich sind Personen auch fähig, die, falls bestehende natürlich bedingte, Kurzsichtigkeit bestimmter Bedürfnisse und deren Befriedigung erfolgreich aufzuschieben bzw. zu sublimieren. Anstatt des Modells einer sich um die Integration gegenläufiger Interessen selbst bemühenen einheitlichen Person setzt Ainslie das Modell des Selbst als Population verschiedener Interessen. Die Integration dieser verschiedenen Interessen erfolgt über die Knappheit an verfügbaren Ressourcen bedingt durch den Umstand, dass diese Interessen an ein und demselben Lebewesen zugleich auftreten und nicht alle gleichzeitig erfüllt werden können.

„In this model the person comprises a variety of interests – groups of reward-seeking processes – that have been shaped by a natural selection process. The selection is based on hyperbolically discounted rewards, and the interests compete strategically for selection in a marketplace that is almost as free as the natural selection of species.“ (Ainslie 2004, 62)

Drei Fragen möchte ich nun an dieses Konzept stellen: a) Inwiefern ist für Ainslie das Konzept von Personen als Populationen verschiedener Interessen notwendig, um irrationalen Phänomene

---

<sup>225</sup> „Dieser Prozess – das aufgrund eines endlichen Vorrats an Äußerungsmitteln nötige Aushandeln der Machtverhältnisse – ist vielleicht schon alles, was eine Person zu einer Einheit macht.“ (Ainslie 2005a, 158)

in die Nutzentheorie zu integrieren? b) Wie wird das Phänomen der intertemporalen Verhandlung zwischen verschiedenen Interessen verstanden? c) Inwiefern ist die Reduktion des Personenbegriffs auf den Begriff einer Population verschiedener Interessen überhaupt plausibel?

a) Die Argumentation von Ainslie lässt sich folgendermaßen als *reductio* rekonstruieren:

- (P1) Personen diskontieren grundsätzlich hyperbolisch, nicht exponentiell. (These der hyperbolischen Diskontierung)
- (P2) Personen maximieren den hyperbolisch diskontierten Nutzen, den sie in Aussicht haben. (These der Nutzenmaximierung unter der Bedingung hyperbolischer Diskontierung)
- (P3) Personen handeln aber manchmal so, als ob sie Nutzen exponentiell diskontieren. (These der Möglichkeit rationalen, auf langfristige Interessen ausgerichteten Handelns)
- (P4) Exponentielle Diskontierung maximiert tatsächlich den Nutzen. (These von der pragmatischen Überlegenheit exponentiellen Diskontierens)
- (P5) Wenn exponentielle Diskontierung tatsächlich den Nutzen maximiert, und Personen manchmal handeln können, als ob sie exponentiell diskontieren, dann würden Personen immer exponentiell diskontieren, um ihren Nutzen zu maximieren.
- (K) Personen diskontieren grundsätzlich exponentiell, nicht hyperbolisch.

Kern des Argumentes ist, dass wenn Personen fähig wären, exponentiell zu diskontieren, sie es immer tun würden, da diese Art der Diskontierung tatsächlich ihren Nutzen maximierte.

„Wenn jemand es lernen könnte, die Diskontkurve, mit der er zur Welt gekommen ist, zu ändern, bestünde stets ein starker Anreiz, dies soweit wie möglich zu tun: Wenn dir eine weit entfernte Aussicht – etwa ein Weihnachtsgeschenk – jetzt nur 3 Nutzeinheiten wert ist, aber du durch einen Trick dafür sorgen könntest, dass die Sache 5 Nutzeinheiten wert wird, könntest du dich allein durch diesen Trick um 2 Nutzeinheiten reicher machen. Wenn du unmittelbar auf deine Diskontkurven einwirken könntest, wärest du dazu in der Lage, dir sozusagen deine eigenen Belohnungen zu schaffen. Man würde damit rechnen, dass jeder diesen Trick ständig und so stark wie möglich ausnutzt. [...] Der springende Punkt ist, dass wir Personen – und übrigens auch Tiere – untersuchen, nachdem bereits jede Abflachung der Kurve erfolgt ist, zu der sie fähig sind. Die Hyperbel, auf die wir stoßen, ist das, was diesen Lernvorgang überdauert hat. Der Trick, durch den man sich so verhält, als wäre die Kurve exponentiell, kann keine Modifikation der Basiskurve sein, sondern muss Mühe kosten und kann nur hin und wieder zum Einsatz kommen.“ (Ainslie 2005a, 151f.)

Es braucht demnach eine These, die zeigt, weshalb Personen unter der Annahme der Nutzenmaximierung<sup>226</sup> und hyperbolischen Diskontierung nicht ständig rational handeln, auch wenn sie fähig sind, zumindest zeitweise exponentiell zu diskontieren und entsprechend zu handeln. Dies soll die These der Zersplitterung des Selbst in voneinander unabhängige Interes-

---

<sup>226</sup> Diese These fungiert als eine der Prämissen von Ainslie und wird demnach nicht hinterfragt. Sein Ziel ist es, eine Lösung zur Beschreibung und Erklärung irrationalen Verhaltens unter der Bedingung der Nutzenmaximierung zu geben.

sensgruppen und die Annahme der *strategischen* Interaktion dieser Interessen untereinander am Markplatz genannt ‚Person‘ leisten. Fraglich ist, ob diese Selbstbindungsprozesse innerhalb der Nutzentheorie über das Konzept zersplitterter Personen integriert werden können. Das vorige Argument zeigt zumindest, dass eine zusätzliche These nötig ist, um sowohl irrationales als auch rationales Verhalten unter der Bedingung der Nutzenmaximierung beschreiben und erklären zu können, und sei es die Einschränkung der These der Nutzenmaximierung unter bestimmten Bedingungen.

b) Parallel zu Ainslies revisionärem Personenkonzept geht seine Interpretation des Willens *als intertemporale Verhandlungssituation*. Durch das strategische Moment können sich einzelne Interessen, wenn sie sich gegenwärtig gegenüber andere Interessen durchsetzen, über ihre gegenwärtige Dominanz hinaus dauernden Einfluss auf das Verhalten bzw. Handeln der Person sichern. So könnte ein langfristiges Interesse an einer ausgewogenen Pension die Person dazu veranlassen, einen Vertrag zur Pensionsvorsorge zu schließen, welcher Sanktionen für den Fall vorsieht, dass nicht pünktlich bezahlt oder der Vertrag vorzeitig gelöst wird. Selbstbindungsprozesse dieser Art sind vielfältig hinsichtlich der Art der involvierten Sanktionsmechanismen – ökonomisch, sozial, symbolisch etc. –, der Gegenstände der Selbstbindung, der Zeitrahmen, der Implementierung usw.<sup>227</sup> Ainslie nennt vier Methoden, welche Interessen nutzen können um zukünftige Entscheidungen der Person zu ihren Gunsten zu beeinflussen:<sup>228</sup> a) äußere Bindung (extrapsychic commitment): Eingrenzung der zukünftig möglichen Optionen; b) Manipulation der Aufmerksamkeit (manipulation of attention); c) Umgang mit Emotionen (preparation of emotion); d) persönliche Regeln (personal rules). Andererseits hören die Interessen nicht auf, miteinander in Konflikten zu stehen und gegeneinander zu konkurrieren. Außerdem können sie die instinktiv angelegte, hyperbolische Weise der Diskontierung nicht dauerhaft ändern.

Zentrale Funktion der Selbstbindung ist die Fähigkeit, persönliche Regeln aufzustellen (d), also Entscheidungen über Einzelfälle als prinzipielle Entscheidungen über Mengen von Einzelentscheidungen aufzufassen, mithin das Fassen von Plänen und das Formen von Prinzipien oder Maximen des Handelns. Nach Ainslie folgt aus der Theorie des hyperbolischen Diskontierens, dass das Zusammenfassen von Entscheidungen über Einzelfälle *motivational* die Effektivität langfristiger Interessen befördert. Es macht demnach einen motivationalen quantitativen Unterschied, ob eine Person eine Entscheidung als reine Einzelentscheidung oder als eine Entscheidung über eine Serie von Entscheidungen auffasst.<sup>229</sup> Nach Ainslie befinden sich

---

<sup>227</sup> Vgl. Elster 2000 zu verschiedenen Mechanismen der Selbstbindung.

<sup>228</sup> Vgl. Ainslie 2004, 73–85.

<sup>229</sup> Vgl. Ainslie 2004, 78–85.

die einzelnen Interessen in der Situation eines wiederholten Gefangenendilemmas, da sie einerseits nur eigeninteressiert sind und andererseits wissen, dass gegenwärtige Entscheidungen Präzedenzfälle für zukünftige Kooperationen darstellen, da die ‚Gefangenen‘ auch wissen, dass sie in Zukunft wieder aufeinander treffen werden. Die Bündelung von Entscheidungen zu Entscheidungen über Serien erfolgt über eine rekursive Funktion von Selbstvoraussagen: Wenn ich heute diesem Interesse nachgebe, dann werde ich bei der nächsten Entscheidung unter denselben Umständen wahrscheinlich wieder nachgeben. Damit wird die Entscheidung für die konkrete Situation verknüpft mit Entscheidungen über analoge Situationen in der Zukunft.

„The bundling phenomenon implies that you will serve your long-range interest if you can obey a *personal rule* to behave alike toward all the members of a category. [...] It also explains how people with fundamentally hyperbolic discount curves may sometimes learn to choose as if their curves were exponential.” (Ainslie 2005b, 641)

Wie gelangen jedoch Personen dazu, Einzelentscheidungen zu Entscheidungsbündel zusammenzufassen, also gegenwärtige Entscheidungen als gute Indikatoren für zukünftige Entscheidungen aufzufassen? Während langfristige Interessen das Aufstellen persönlicher Regeln motivieren, untergraben kurzfristige Interessen diese wiederum durch die Motivation für Ausnahmen. Da keine Regel so präzise bestimmt werden kann, dass sie ohne weiteres auf konkrete Situationen angewendet werden können, bleibt immer ein gewisser Spielraum der Interpretation übrig, den kurzfristige Interessen ausnützen können.<sup>230</sup> Unklar ist, inwiefern die Bündelung von Entscheidungen in einer Situation, die laut Ainslie zusätzliche motivationale Kräfte entfaltet, nicht einfach nur eine Änderung der ursprünglichen Entscheidungssituation darstellt: die Entscheidung, heute Schokolade zu essen oder nur heute zu fasten ist eine andere als die, heute Schokolade zu essen oder meine langjährige Diät einzuhalten. Der mit diesen Optionen verbundene Erwartungsnutzen ist jeweils verschieden, wodurch auch bei exponentieller Diskontierung andere Ergebnisse zu erwarten sind. Aber Ainslie geht es um eine bottom-up Analyse der menschlichen Fähigkeit, quasi-exponentiell zu diskontieren, und dafür ist die Eigenschaft der hyperbolischen Diskontierung, bei summierten Entscheidungen tendenziell langfristige Interessen zu stärken, zentral.

c) Die Ausführungen von Ainslie zu seinem revisionären Personenkonzept sind allesamt höchst metaphorisch und problematisch.<sup>231</sup> Der Vorwurf, der mit dieser Metaphorik verbunden ist, ist

---

<sup>230</sup> „Intertemporal cooperation is most threatened by rationalizations that permit exceptions for the choice at hand, and is most stabilized by finding *bright lines* to serve as criteria for what constitutes cooperation. [...] Long range interests define principles, and short range interests find exceptions.” (Ainslie 2005b, 642) Vgl. diese Mechanismen bei Wolf und Schälike in Kap. III.1.

<sup>231</sup> „The ego isn’t an organ that might sit in a central place like Descartes’s pineal-based soul. It’s a network server, a broker of cooperation among the interests, and, like interests, is itself engendered and shaped by differential reward – specifically, by the long-range reward that comes from better defense against short-range rewards.” (Ainslie 2004, 99)

ein klassischer *Homunkulus*-Einwand:<sup>232</sup> Wie soll die Annahme kleiner, quasi-personaler Einheiten erklären, was die Annahme einer einzigen großen personalen Einheit nicht gelingt? Was heißt es, dass das Selbst keine Einheit ist, sondern der intertemporale Prozess des Aushandelns von Interessenskonflikten auf dem inneren Marktplatz? Was heißt es überhaupt, dass Interessen zur *strategischen* Interaktion fähig sind? Hier werden Begriffe, die auf Personen zugeschnitten sind, auf sub-personale Einheiten angewendet, und dies verlangt eine ausreichende Klärung der Rechtmäßigkeit einer solchen Anwendung. Zuerst zum Problem der fehlenden phänomenalen Ausweisung des Modells in der eigenen, alltäglichen Erfahrung. Ainslie sieht hier drei Einwände, die gegen sein Modell sprechen:

*Erstens*: intertemporale Transaktionen involvieren Bewusstsein dieser Transaktionen, wir sind uns unserer ständigen intertemporalen Transaktionen aber nicht bewusst.<sup>233</sup> Hier meint Ainslie, dass Selbsttäuschungsprozesse uns davon abhalten, die eigenen intertemporalen ‚Geschäfte‘ als solche wahrzunehmen. Personen nehmen eigens aufgestellte Regeln als externe Fakten wahr, was zugleich die Effektivität der Regelbefolgung erhöhe.<sup>234</sup> Den Ausführungen Ainslies nach ist also der Glaube an Werte, Regeln und Normen, die unabhängige Geltung von ihrem faktischen Geglautwerden beanspruchen, selbst nur ein Mechanismus, um die Effektivität der Regelbefolgung zu steigern. Dies deshalb, da selbst gemachte Regeln leichter zu ändern sind als von außen gesetzte oder geltende Regeln.

„Now we can make a guess at why the theory of hyperbolic discounting feels wrong at first: Most people *believe in* the rational value structure implied by exponential discounting, which we experience as consistent choice over time. We hold it out as a property of nature, which every normal person will discover in time. It may be just a quibble to say instead that exponential discounting is the inevitable equilibrium point that people largely learn to adopt in their long-range interest after they get experience bargaining with shorter-range interests. But bargaining is something we *do*. People want to characterize value as something we *discern*, for the very reason – my argument goes – that framing the rule as a belief will make it more stable.”  
(Ainslie 2004, 110)

Für den diese Erklärungen unterliegenden Nonkognitivismus oder Werteneihilismus argumentiert Ainslie nicht. Ob hyperbolische Diskontierung ein zusätzliches Argument in dieser Frage liefert, ist fraglich: Nur weil ein Teil unserer Überzeugungen motiviert ist durch deren Effektivität, bestimmten Interessen zu dienen, und damit in Wahrheit soziale Konstrukte darstellen, heißt dies nicht, dass alle normativen Überzeugungen so beschaffen sind oder das es rational wäre, dies anzunehmen.

---

<sup>232</sup> „Ainslie follows a tradition dating back to Plato when he proposes disassembling this “I” into a group of lesser agents or agencies, and composing the competence (and incompetence) of the whole person out of the interactions of these subpersonal homunculi, which he calls “interests”. There is nothing wrong with this tactic so long as none of the homunculi are *too smart*.” (Griffin & Dennett, 656) Vgl. die Diskussion des Homunkulus-Einwandes gegenüber Sokrates bzw. Platon bei Müller 2009, 104–108.

<sup>233</sup> Vgl. Ainslie 2004, ch. 7.1.

<sup>234</sup> Vgl. Ainslie 2004, 106ff.

*Zweitens:* ständige Transaktionen benötigen andauernde Aufmerksamkeit um sie durchzuführen, was intellektuell überlastend wäre.<sup>235</sup> Hier meint Ainslie, dass viele der Abwägungen für diese Transaktionen sich in Form von Gewohnheiten niederschlagen und daher nicht ständig neu hinterfragt werden. An diese Frage schließt natürlich die weitere an, weshalb wir manche Entscheidungen gewohnheitsmäßig vollziehen, andere hingegen neu überdenken und neue Abwägungen durchführen. Im Grunde bleibt der Prozess der Abwägung von Interessen, der intertemporalen Transaktion, mysteriös. Innerhalb der Theorie soll es keine eigene Instanz der Abwägung geben, sondern die Illusion einer solchen Instanz ist lediglich das Produkt einer fragilen Stabilität von Kooperationen von Einzelinteressen, die unter den Bedingungen von Nutzenmaximierung, hyperbolischer Diskontierung und der Situation von sich ständig wiederholenden Gefangenendilemmata entsteht.

„Contradictory reward-getting processes can in effect bargain with each other, and stable preferences can be established by the perception of recurrent choices as test cases (precedents) in recurrent *intertemporal* prisoner’s dilemmas.” (Ainslie 2005b, 635)

Sogenannte Ego-Funktionen bzw. die Ausbildung eines Willens sind das Produkt der Anpassungsleistungen der Interessen an die jeweilige Marktlage innerhalb des Individuums sowie die Effektivität des Willens in der Vergangenheit, für Ausgleich zwischen den miteinander konfligierenden Interessen zu sorgen.

*Drittens:* das Modell sieht vor, dass jede Entscheidung rekursiv Bezug auf alle zukünftigen Entscheidungen nimmt, aber nur wenige Entscheidungen sind solche Alles-oder-nichts-Entscheidungen.<sup>236</sup> Ainslie entgegnet, dass die meisten dieser intertemporalen Transaktionen klein sind und nur wenige Entscheidungen den Charakter von schicksalhaften Entscheidungen über die gesamte Zukunft haben.

Insgesamt bleibt problematisch, dass nicht klar ist, was einzelne Interessen zur strategischen Interaktion und Kooperation befähigt. Hier ergeben sich zwei Einwände: a) Offensichtlich bedarf es beträchtlicher kognitiver Fähigkeiten, um strategisch und auf die Zukunft gerichtet auf dem intertemporalen Markt der Interessen zu bestehen. Es bleibt völlig unklar, wie diese kognitiven Fähigkeiten mit den quasi-personalen Einheiten genannt Interessen zusammenhängen. b) Selbst unter diesen Annahmen ist es unklar, ob unter diesen Bedingungen der Anreiz für kurzfristige Interessen groß genug ist, sich auf intertemporale Kooperationen mit anderen Interessen einzulassen. Angenommen, ein kurzfristiges Interesse erlangt kurzzeitige motivationale Dominanz über die anderen. Weshalb sollte dieses Interesse auf ihre gegenwärtige Befriedigung zugunsten anderer Interessen verzichten? Das Interesse würde nur dann darauf verzichten, wenn

---

<sup>235</sup> Vgl. Ainslie 2004, ch. 7.2.

<sup>236</sup> Vgl. Ainslie 2004, Kap. 7.3.

es nicht nur kurzfristig wäre, sondern auch ein Interesse daran ist, auch in Zukunft befriedigt zu werden. Das Problem, dass Interessen *prima facie* keine geeigneten Entitäten für strategische Interaktionen sind, ist hartnäckig. Hier müsste Ainslie zeigen, wie die verschiedenen Interessen von den kognitiven Ressourcen Gebrauch machen können, die für diese Art der strategischen Interaktion nötig sind, ohne dabei zu echten personalen Einheiten zu werden, wodurch das Einheitsproblem schlagend wird und der Homunkulus-Einwand gerechtfertigt wäre. Fraglich ist auch, ob das Modell des Selbst als einer Population von verschiedenen Interessen tatsächlich notwendig ist, um die irrationalen Phänomene auch unter nutzentheoretischen Maximen adäquat zu beschreiben. Die Rolle des Wissens über sich selbst, über das eine Person verfügt, und das in expliziter wie impliziter Form vorliegt, das sich in Voraussagen über den Einfluss von Entscheidungen auf die zukünftige Entscheidungsfähigkeit ausdrückt, sollte auch unter der Annahme eines traditionellen Modells einer handlungsfähigen, einheitlichen Person wichtig sein. Diesen Einwand formuliert Luca Ferrero folgendermaßen:

„I am unpersuaded by Ainslie’s central claim that the will is the product of transtemporal bargaining among successively dominant, transient interests analogous to the emergence of cooperation in a repeated Prisoner’s Dilemma [...]. It is questionable that we could make sense of the parties involved in this bargaining as truly separate sources of agency. And even if we could, it is hard to see how cooperation could emerge out of the interactions between these parties. It is doubtful, therefore, that the will can be understood as a genuine *interpersonal* phenomenon.“ (Ferrero 2005, 654)

Für eine adäquate Konzeption des Willens sind die Fähigkeit zu und die Auswirkungen von Selbstvorhersagen über das eigene Handeln und die zukünftige Handlungsfähigkeit sicherlich ein wichtiger Bestandteil. Dies ist jedoch auch keine neue Entdeckung, sondern implizit in jeder Handlungstheorie angelegt, die meint, dass die Intention, etwas zu tun, zumindest impliziert, dass die Person selbst glaubt, dass es ihr nicht vollkommen unmöglich ist, die vorgestellte Handlung zu vollziehen. Ferrero stellt daher die berechtigte Vermutung an, dass das Element der Voraussage besser *intrapersonal* modelliert wird, statt *interpersonal* als Verhandlungsergebnis quasi-personaler Einheiten. Für eine Person, die um ihre zeitliche Ausdehnung weiß, sowie über ihre Tendenz der temporalen Kurzsichtigkeit bei Konflikten zwischen kurz- und langfristigen Interessen, kommt der Ausbildung eines Willens eine ebenso funktionale Rolle wie in dem Modell von Ainslie, jedoch ohne die strittige Reduktion des Personenbegriffs.<sup>237</sup>

---

<sup>237</sup> Vgl. die Kritik an Konzeptionen mit sukzessiven ‚Selbsten‘ von Jon Elster: „Barring pathological cases (which I have not discussed here) we ought not to take the notion of ‘several selves’ very literally. In general, we are dealing with exactly *one* person – neither more nor less. That person may have some cognitive coordination problems, and some motivational conflicts, but it is *his* job to sort them out. They do not sort themselves out in an inner arena where several homunculi struggle to get the upper hand.“ (Elster 1987, 30f.)

### III.3 Semantische Theorien und Sprechakttheorien

Zu dieser Gruppe gehören Theorien, die vornehmlich auf Basis semantischer und sprechakttheoretischer Überlegungen über evaluative und normative Urteile Konsequenzen für das Phänomen der Willensschwäche ableiten. Ich diskutiere hierbei die präskriptivistische Analyse von Richard Hare, danach eine sprechakttheoretisch motivierte Kritik von Robert Dunn an dieser Analyse.

#### III.3.1 Richard Hare: Präskriptivismus und Willensschwäche

Vor allem in drei Veröffentlichungen – *The Language of Morals* (1952), *Freedom and Reason* (1963) und *Moral Thinking* (1981) – hat Richard Hare seine metaethische Theorie über die Semantik moralischer Ausdrücke, seinen universellen *Präskriptivismus*, ausformuliert. Relevant ist diese Theorie hier wegen seiner darin enthaltenen Auseinandersetzung mit dem Phänomen der Willensschwäche. Nach Hare folgt aus dem präskriptiven und universalistischen Charakter von evaluativen Ausdrücken, dass synchrone Willensschwäche unmöglich ist bzw. die betreffende Person physisch und/oder psychisch in diesen Fällen nicht anders handeln konnte, als sie es getan hat. Die Bedeutung von normativen Ausdrücken impliziere demnach, dass Personen immer das tun, was sie selbst für das Beste halten. Das Phänomen der Willensschwäche interessiert Hare dabei nur als mögliches Gegenbeispiel für seinen Präskriptivismus.<sup>238</sup> Außerdem geht es Hare nicht um Willensschwäche allgemein, sondern er beschränkt sich auf Willensschwäche als moralische Schwäche (moral weakness oder auch backsliding), als Handeln wider ein Moralurteil, dem die Person aufrichtig zustimmt.

Der Präskriptivismus, als eine Spielart des Nonkognitivismus, interpretiert die Semantik von evaluativen und vor allem moralischen Ausdrücken anhand ihrer Funktion, Handlungen anzuleiten.<sup>239</sup> Evaluative Ausdrücke haben sowohl eine deskriptive (phrastic component) als auch eine präskriptive Komponente (neustic component), wobei die letztere entscheidend ist für evaluative Ausdrücke.<sup>240</sup> Da die Funktion von evaluativen Urteilen darin bestehe, Handlungen anzuleiten, interpretiert Hare das evaluative Element als Imperativ. Eine Person, die einem

---

<sup>238</sup> Vgl. Spitzley 1992, 125–164; Taylor 1980; Pears 1984, 223ff. Zur Verteidigung von Hare vgl. Silverstein 1970.

<sup>239</sup> Aber: Hare selbst sieht seine Theorie nicht als eine Form von Nonkognitivismus an, da er meint, dass seine Theorie Kriterien für die Wahrheit bzw. Falschheit von moralischen Urteilen liefere. Im Vergleich zu klassischen Versionen von nonkognitivistischen Theorien wie diejenigen von Alfred Ayer oder Charles Stevenson ist Hares Präskriptivismus sicherlich ‚kognitiver‘. Doch es lässt sich zweierlei einwenden: 1) Das präskriptive und somit eigentlich normative Element ist weder wahr noch falsch; 2) Die Wahrheit evaluativer Ausdrücke besteht in ihrer Konsistenz mit einem herrschenden Standard, wird dementsprechend nicht realistisch als Korrespondenz mit evaluativen Eigenschaften gedeutet. Vgl. Hare 1993.

<sup>240</sup> Vgl. Hare 1952, 18. „Prescriptivists hold that moral statements are *expressions of volition* [...]“ (Hare 1998)

evaluativen Urteil zustimme, sei analog zu sehen zu jemandem, der einem Imperativ an sich selbst zustimme:

„I propose to say that the test, whether someone is using the judgement ‘I ought to do X’ as a value-judgement or not is, ‘Does he or does he not recognize that if he assents to the judgement, he must also assent to the command “Let me do X?”’” (Hare 1952, 168f.)

Dabei versteht Hare dies als Definition dessen, was es heißt, einen Ausdruck evaluativ zu gebrauchen bzw. einen vorliegenden evaluativen Ausdruck im vollen Sinn zu verwenden. Nonkognitivistisch ist diese Theorie, da Imperative keinen Wahrheitswert haben, d.h. weder wahr noch falsch sein können. Abgeschwächt wird der Nonkognitivismus über das deskriptive Element und die damit verbundene Universalisierbarkeit von moralischen Urteilen:<sup>241</sup> Moralische Urteile haben eine inhärente Tendenz zur Universalisierung, d.h. zur Anwendung auf ähnliche Situationen, in denen das ursprüngliche Moralurteil gefällt wurde. Diese Verbindung von deskriptiven und präskriptiven Elementen soll sowohl die rationale Argumentation für ethische Urteile ermöglichen als auch deren nicht-natürlichen Charakter verständlich machen.<sup>242</sup> Da Hare außerdem davon ausgeht, dass zwischen Sollen und Können eine enge Verbindung besteht, insofern eine Person, wenn sie etwas tun soll, dies auch tun kann, folgt für ihn, dass Personen, die einem moralischen Urteil zustimmen, und damit einem Imperativ sich selbst gegenüber zustimmen, entsprechend ihrem Urteil handeln *oder* psychisch bzw. physisch unfähig sind, entsprechend zu handeln.

„It is a tautology to say that we cannot sincerely assent to a second-person command addressed to ourselves, and *at the same time* not perform it, if now is the occasion for performing it and it is in our (physical and psychological) power to do so. Similarly, it is a tautology to say that we cannot sincerely assent to a statement, and *at the same time* not believe it. Thus we may characterize provisionally the difference between statements and commands by saying that, whereas sincerely assenting to the former involves *believing* something, sincerely assenting to the latter involves (on the appropriate occasion and if it is within our power), *doing* something.” (Hare 1952, 20)

Willensschwäche fungiert somit auch hier als mögliches Gegenbeispiel:

„The ethical theory which has been briefly set out in the preceding chapters is a type of prescriptivism, in that it maintains that it is one of the characteristics of moral terms, and one which is a sufficiently essential characteristic for us to call it part of the meaning of these terms, that judgements containing them are, as typically used, intended as guides to conduct. Now there is one objection to all kinds of prescriptivism which is so commonly made, and is of such intrinsic interest, that it requires a chapter to itself. This is the objection that, if moral judgements were prescriptive, then it would be impossible to accept some moral judgement and yet act contrary to it.” (Hare 1963, 67)

Wenn Moralurteile Imperative implizieren, und diese wiederum Handlungen implizieren, dann können Personen nicht willensschwach handeln, wenn dies hieße, wider ein Moralurteil zu

---

<sup>241</sup> Das deskriptive Bedeutungselement evaluativer Ausdrücke liefert Gründe für die evaluativen Urteile. Vgl. Hare 1963, 21; Huemer 2005, 25–30; Spitzley 1992, Kap. 4.

<sup>242</sup> Für die Kritik von Hare am ethischen Naturalismus vgl. Hare 1952, 79–93.

handeln, dem die Person im vollen Sinne (präskriptiv) und aufrichtig zustimme.<sup>243</sup> Die Qualifikation ‚im vollen Sinne‘ ist für Hare nötig, da Personen evaluative Ausdrücke auch auf rein deskriptive Weise oder nicht universell verwenden können (inverted comma sense) und damit gerade nicht streng evaluativ.

Die Argumentation in Bezug auf Willensschwäche kann demnach folgendermaßen rekonstruiert werden:

- (P1) Moralische Urteile sind handlungsanleitend.
- (P2) Wenn moralische Urteile handlungsanleitend sind, dann impliziert die aufrichtige Zustimmung einer Person zu einem Moralurteil die Zustimmung zu einem Imperativ, entsprechend dem Urteil zu handeln. (These der Präskriptivität evaluativer Ausdrücke)
- (P3) Imperative, die an sich selbst gerichtet sind, werden befolgt, oder die Person ist physisch oder psychisch unfähig, entsprechend zu handeln. (These über an sich selbst gerichtete Imperative)
- (K) Also: Ein Moralurteil, dem eine Person aufrichtig zustimmt, wird auch umgesetzt, wenn die Person physisch und psychisch dazu in der Lage ist.

Für vermeintliche Willensschwäche ergeben sich innerhalb der Theorie folgende Möglichkeiten:

a) die Person stimmt ihrem Moralurteil gar nicht aufrichtig zu, d.h. sie fällt das Moralurteil nur in einem abgeschwächten, nicht evaluativen, d.h. bei Hare nicht präskriptiven Sinn (inverted comma sense);<sup>244</sup> b) das Moralurteil ist nicht universell, so dass es die eigene Person und das konkrete eigene Handeln nicht umfasst – es lässt also die gegenwärtige Situation als Ausnahme zu; c) die Person ist psychisch oder physisch nicht in der Lage, entsprechend ihrem Moralurteil zu handeln. Der typische Fall von Willensschwäche für Hare ist c), also der einer Person, die etwas tun soll,

---

<sup>243</sup> „But to guide choices or actions, a moral judgement has to be such that if a person assents to it, he must assent to some imperative sentence derivable from it; in other words, if a person does not assent to some such imperative sentence, that is knock-down evidence that he does not assent to the moral judgement in an evaluative sense – though of course he may assent to it in some other sense [...]. This is true by my definition of the word ‘evaluative’. But to say this is to say that if he professes to assent to the moral judgement, but does not assent to the imperative, he must have misunderstood the moral judgement (by taking it to be non-evaluative, though the speaker intended it to be evaluative). We are therefore clearly entitled to say that the moral judgement entails the imperative; for to say that one judgement entails another is simply to say that you cannot assent to the first and dissent from the second unless you have misunderstood one or the other; and this ‘cannot’ is a logical ‘cannot’ – if someone assents to the first and not to the second, this is in itself a sufficient criterion for saying that he has misunderstood the meaning of one or the other. Thus to say that moral judgements guide actions, and to say that they entail imperatives, comes to much the same thing.” (Hare 1952, 171f.)

<sup>244</sup> Die Person urteilt z.B. nur so, wie sie denkt, dass *andere* Personen moralisch oder generell evaluativ über den Gegenstand urteilen würden. Das Urteil wäre demnach nur ein konventionelles Urteil der Art ‚*Man* sollte X tun/unterlassen‘ und nicht ein moralisches Urteil der Art ‚*Ich* sollte X tun/unterlassen‘. Zu ‚inverted comma sense‘ vgl. Hare 1963; Brink 1989, 46ff.

dies aber nicht tun kann.<sup>245</sup> Dies ist nachvollziehbar, da c) die einzige Option ist, in der die willensschwache Person tatsächlich einen Konflikt durchlebt, da in den beiden anderen Fällen die Person gar nicht wirklich urteilt, dass sie anders handeln sollte, als sie es tut.

In *Moral Thinking* fügt Hare dieser Liste an Möglichkeiten noch eine weitere hinzu. Dort unterscheidet er zwischen zwei Ebenen des moralischen Denkens, einer *intuitiven* und einer *kritischen*.<sup>246</sup> Auf der intuitiven Ebene denken wir im Alltag, wenn wir unsere Handlungen in konkreten Situationen an Moralprinzipien ausrichten. Diese nennt Hare *prima facie* Moralprinzipien, die im Gegensatz zu denen auf der kritischen Ebene Ausnahmen zulassen, d.h. übertrumpft werden können. Auf die kritische Ebene greifen wir zurück, wenn in bestimmten Situationen unsere *prima facie* Moralprinzipien einander widersprechen, d.h. verschiedene Handlungen verlangen, und wir diese Konflikte lösen und dabei unsere *prima facie* Moralprinzipien der intuitiven Ebene verfeinern. Diese Konzeption lässt eine weitere Art von Willensschwäche zu: Da *prima facie* Moralprinzipien prinzipiell so beschaffen sind, dass sie übertrumpfbar sind, können sie auch von nichtmoralischen Prinzipien in einzelnen Situation übertrumpft werden.<sup>247</sup>

„That is why the whole problem becomes clearer when one sees the kind of conflict which we call weakness of will as just one example of conflicts between prescriptions. What happens when I decide that I ought to break a promise in order not to disappoint my Australian friend of his tour of Oxford has quite close affinities with what happens when I decide to break one in order not to disappoint my own appetites.” (Hare 1981, 60)

Was Hare für diese Beschreibung von Willensschwäche als Konflikt zwischen moralischen *prima facie* Pflichten und nicht-moralischen *prima facie* Pflichten voraussetzt ist, dass Begierden (appetites) solchen Präskriptionen zumindest korrespondieren.<sup>248</sup> Kritik an dieser Theorie kann auf mehrere Weise geübt werden:<sup>249</sup>

a) Kritik an der generellen Präskriptivismusthese (P2): Weshalb sind evaluative Urteile als versteckte Imperative, und nicht einfach als Indikative zu interpretieren? Dieser Punkt folgt sogleich in der Besprechung von Robert Dunns Kritik an Hare.

b) Kritik an der Imperativthese (P3): Selbst wenn evaluative Urteile präskriptiv gedeutet werden, ist nicht klar, weshalb nicht wider einem solch aufrichtig gefällten Urteil gehandelt werden

---

<sup>245</sup> „Now we shall see that typical cases of 'moral weakness' are cases where a man *cannot* do what he thinks he ought [...]” (Hare 1963, 68) Für Kritik am Begriff des ‚psychischen Zwanges‘ bei Hare vgl. Spitzley 1992, Kap. 4.2.

<sup>246</sup> Vgl. Hare 1981, 25.

<sup>247</sup> Vgl. Hare 1981, 59f.

<sup>248</sup> Hare drückt sich nicht eindeutig aus, ob diese Präskriptionen *identisch* sind mit Wünschen, ob sie Wünschen korrespondieren oder diese nur sprachlich ausdrücken. Vgl. Spitzley 1992, 151f. Vgl. Kap. II.2 über Davidson, für den Wünsche mit *prima facie* evaluativen Urteilen korrespondieren.

<sup>249</sup> Vgl. Taylor 1980; Spitzley 1992, Kap. 4; Frankena 1988.

könnte.<sup>250</sup> Ist es tatsächlich so, dass wir Imperativen stets folgen, wenn wir sie an uns selbst richten? Wie Robert Dunn bemerkt, sind Imperative normalerweise mit Intentionen verbunden, dass *andere* Personen etwas tun oder unterlassen, nicht mit Intentionen, selbst etwas zu tun oder zu unterlassen.<sup>251</sup> Bei Imperativen, die wir an andere richten, ist es nun offensichtlich so, dass wir nicht notwendigerweise voraussetzen, dass Personen diesen folgen, selbst wenn sie psychisch und physisch dazu in der Lage sind. Wenn es also der Fall wäre, dass Intentionen, etwas zu tun, mit Imperativen an sich selbst ausdrückbar wären, könnte also umgekehrt diese Tatsache als Indiz dafür gedeutet werden, dass wir uns nicht sicher sind, ob wir diesen Imperativen folgen oder nicht. Es ist unklar, weshalb jemand nicht einen Imperativ an sich selbst richten könnte, und zwar in einem vollen Sinn und aufrichtig, und die sich selbst vorgeschriebene Handlung dennoch nicht vollzieht.<sup>252</sup> Augenscheinlich wird diese Möglichkeit daran, dass folgende Sätze offensichtlich nicht sinnlos sind und sich die Sprecherin solcher Aussagen nicht selbst widerspricht wie jemand, der sagt ‚Heute regnet es, aber ich glaube nicht, dass es regnet‘:

‚Ich soll X tun, aber ich beabsichtige nicht, X zu tun.‘

‚Ich denke, dass ich jetzt X tun soll, aber ich beabsichtige nicht, X jetzt zu tun.‘

‚Ich bin nach Abwägung aller Möglichkeiten der Meinung, dass ich am besten X tun sollte. Aber ich habe mich noch nicht dazu entschlossen, X zu tun, und werde mich wahrscheinlich auch nicht dazu entschließen.‘

‚Ich befehle mir, X zu tun, aber ich beabsichtige nicht, X zu tun.‘

---

<sup>250</sup> Vgl. die Kritik bei Pears 1984, 242ff. Nach David Pears folgt aus Hares Theorie nur, dass die Person, die an sich selbst einen Imperativ richtet, die Intention hat, dass sie eine *weitere* Intention ausbildet, entsprechend dem Befehl zu handeln, nicht, dass sie die Intention hat, entsprechend zu handeln. Es bleibt demnach ein Übergang von der Intention, eine entsprechende Intention zu haben, und diese Intention tatsächlich zu haben. „What happens when the guest very much wants a third drink but tells himself that it is best to stop at two and really means it? According to this theory, he intends himself to stop at two and, therefore, intends to stop at two. However, though the smooth inference from the first intention to the second one would be acceptable in a case where all the agent’s desires pointed in the same direction, it is questionable in a case of conflict. For if the conflict is at all severe, he will not intend to stop at two drinks and that is precisely why he will tell himself that it is best to stop at two with the intention that, as a result, he will form the intention to stop at two and then stop at two.” (Pears 1984, 243) Diese Interpretation würde dennoch lediglich höchstens exekutive Willensschwäche zulassen, da hier ein Konflikt von Intentionen verschiedener Stufen, und nicht einer zwischen besserem Urteil und Intention vorliegt.

<sup>251</sup> Der Einwand sprechakttheoretisch bei Dunn: „Hare contends that (first-person singular) imperatives are suitable sentences for expressing in words what we are thinking when we have volitions *to act*. The crucial issue here is whether directive illocutions, the illocutions with which imperatives are semantically associated, ever count as expressing such volitions. The answer to this, I contend, is no. For directive illocutions are essentially *other-oriented* and hence, always count only as expressions of intentions or desires that *others* should act in certain ways [...]” (Dunn 1987, 65f.)

<sup>252</sup> Immanuel Kant, auf den sich Hare gerne beruft, interpretiert moralische Normen und Prinzipien auch als Imperative, aber gerade deswegen, da die Natur des Wesens, an den diese Imperative gerichtet sind – Menschen als *teilweise* Naturwesen – derart ist, dass sie eben *nicht* immer das tun, was sie als moralisch richtig einsehen: Das Moralgesetz als oberstes Moralprinzip tritt deswegen in Gestalt des kategorischen Imperatives auf. Vgl. Kant 1785, 414.

Die einfachste Möglichkeit, diese Sätze verständlich zu machen, ist anzunehmen, dass Intentionen, etwas zu tun, nicht mit Imperativen, die an sich selbst gerichtet sind, ausdrückbar sind. Die übliche Art, Intentionen auszudrücken ist über *kommissive* Sprechakte (schwören, versprechen, drohen usw.), und diese sind von *direktiven* Sprechakten (auffordern, bitten, befehlen usw.) zu unterscheiden.

Dieser Einwand lässt sich auch handlungstheoretisch ausdrücken: Weshalb sollen an sich selbst gerichtete Imperative identisch sein mit Intentionen, entsprechend zu handeln, oder weshalb sollten solche Intentionen aus diesen notwendigerweise folgen, wenn die Person psychisch und physisch in der Lage ist, entsprechend zu handeln? Die Annahme einer solchen engen Verbindung verlangt zumindest ein unabhängiges Argument, das Hare nicht liefert. Für Hare besteht die Funktion von Moralarurteilen darin, das Handeln anzuleiten: Moralarurteile geben die Antwort auf die Frage, was wir in einer gegebenen Situation tun sollen. Aber: selbst wenn dies die Funktion von Moralarurteilen ist, bleibt die Frage bestehen, ob wir, wenn wir wissen, was wir tun sollen, auch schon die Frage geklärt haben, *ob* wir tatsächlich das tun, was wir tun sollen. Wenn ein Imperativ die Frage für eine Person beantwortet, was sie tun soll, steht noch nicht fest, dass sie tatsächlich tut, was der Imperativ ihr vorschreibt, selbst wenn der Imperativ von ihr selbst stammt. Imperative sind nicht die richtige Art von Antwort auf die Frage, ob tatsächlich eine Absicht besteht, etwas zu tun, sondern nur darauf, *wie* gehandelt werden sollte, *falls* gehandelt wird. Im Fall von Imperativen, die an andere gerichtet sind, ist dies hinreichend klar: Niemand würde behaupten, dass mein Problem, was ich tun soll, bereits hinreichend gelöst ist, wenn ich den Befehl erhalte, dieses oder jenes zu tun. Denn die Frage, *ob* ich dem Befehl folge leiste oder nicht, ist mit der aufrichtigen Äußerung des Imperativs seitens der anderen Person noch nicht beantwortet. Weshalb sollte diese Charakteristik sich ändern in dem Fall, in dem ich mir selbst einen Befehl gebe? Richtige Antworten auf die Frage, *ob* ich handle, sind sprachlich gesehen kommissive Sprechakte, in denen sich eine Person auf Handlungen festlegt, bzw. die ihnen entsprechenden *Intentionen* als Bedingung für die Aufrichtigkeit der entsprechenden Äußerungen zu formen. Die von Hare als selbstverständlich vorausgesetzte These, dass Moralarurteile die Funktion haben, Handlungen anzuleiten, ist demnach nicht so einfach und eindeutig zu verstehen, wie sein Präskriptivismus es vermuten lässt.

### **III.3.2 Robert Dunns sprechakttheoretische Kritik**

In seinem Buch *The possibility of weakness of will* (1987) verteidigt Robert Dunn generell die Möglichkeit von Willensschwäche, nämlich verstanden als Handeln wider ein gegenwärtiges, besseres

Urteil über das eigene Handeln.<sup>253</sup> Willensschwäche dieser Form ist möglich, da gegenwärtige evaluative Urteile über das eigene Handeln ausschließlich Ausdruck theoretischen Denkens sein können und nicht notwendigerweise mit volitionalen Akten (Intentionen oder Wünschen), entsprechend zu handeln, verbunden sein müssen oder selbst teilweise oder ganz aus solchen Akte bestehen.

Dunn charakterisiert Hares Theorie als Form eines Volitionalismus evaluativen Denkens, dem gemäß evaluatives Denken selbst teilweise aus volitionalem Denken bestehe. Generell verwische dabei der Volitionalismus den Unterschied zwischen Evaluation (Bewertung) und Wertschätzung:

„Volitionism, it seems to me, seduces us when we lose sight of the difference between *(e)valuating* and *valuing*. What we *value* is what matters or is important to us. That is to say, it is only *valuing* that implies the presence of volitional thinking. Whereas *(e)valuating* something is *merely* thinking it to be of some value, *valuing* it is *both* thinking it to be of value *and* having it (for the reason that it has such value) as an object of volition.” (Dunn 1987, 13f.)

Die Kritik an Hare besteht dabei aus drei Schritten:

1) Evaluative Ausdrücke implizieren kein imperatives Bedeutungselement.<sup>254</sup> Damit weist Dunn den Anspruch von Hare zurück, das präskriptive Element evaluativer Ausdrücke mittels Imperative auszudrücken.<sup>255</sup> Evaluative Ausdrücke implizieren nicht notwendigerweise direkte illokutionäre Akte sondern sind reine Assertionen bzw. indikative Äußerungen.

„If evaluative terms like *good*, *right*, *ought*, and so on, build an imperative element into the meaning of the categorical affirmatives that contain them, evaluative illocutions performed in uttering such categorical affirmatives necessarily will be constituted, at least in part, by directive illocutions. But such evaluative illocutions can be entirely dissociated from directive illocutionary force; therefore prescriptivism, even in its sophisticated restatement, is finally untenable.” (Dunn 1987, 44)

Dunn behauptet nicht, dass evaluative Ausdrücke niemals mit direktiven illokutionären Sprechakten verknüpft sind, sondern nur, dass keinerlei notwendige, durch den Sinn evaluativer Ausdrücke vorgegebene, derartige Verbindung existiert. Das zeige sowohl die syntaktische Struktur solcher Ausdrücke<sup>256</sup> als auch Fälle, in denen evaluative Äußerungen ohne direkte Illokution verständlich geäußert werden können, wie z.B.:

„You ask me what I think is the right thing for Fred to do, and I say: “Well, I’ve no doubts about it. The right thing for Fred to do is to leave at once” – adding – “But that’s between you and me. You wouldn’t catch me dead even suggesting to someone what to do in circumstances like this.” (Dunn 1987, 51)

---

<sup>253</sup> „Is it possible for an agent to be weak-willed in what he does? That is to say, can an agent knowingly and intentionally act against his full-fledged all-out summary better judgement, or judgement about what is right, or some such, when he is free to act in accordance with it, has the relevant know-how, and has present to mind the judgement that now is the time to act? Apparently so. Pre-reflectively, such weakness of will would seem a familiar, everyday phenomenon.” (Dunn 1987, 1)

<sup>254</sup> Vgl. Dunn 1987, ch. 2.

<sup>255</sup> Zu weiterer Kritik am Nonkognitivismus generell vgl. Huemer 2005, ch. 2.

<sup>256</sup> Vgl. Dunn 1987, 44–48.

In diesem Fall liegt eine evaluative Äußerung vor mit einer zusätzlichen klaren Distanzierung seitens des Sprechers von jeder damit verbundenen direktiven illokutionären Absicht.<sup>257</sup> Bei Fällen dieser Art wäre es auch schwierig für Hare ein Rückzugsargument anzubringen der Art, dass der Sprecher hier ‚richtig‘ nur in einem abgeschwächten (inverted comma sense) verwendet hätte.

Auf ein schwieriges Problem, evaluative Äußerungen als zumindest teilweise präskriptive Äußerungen zu verstehen, verweist zudem das sogenannte Frege-Geach Problem.<sup>258</sup> Gottlob Frege war bedacht darauf, die behauptende Kraft, mit der ein Satz geäußert wird, von der Bedeutung und dem Sinn des Satzes zu unterscheiden. Peter Geach hat auf diese Einsicht zurückgegriffen und daraus ein Argument gegen die nonkognitivistische Interpretation, vor allem emotivistischer Machart, gezogen. Der Einwand gegen den Nonkognitivismus in diesem Zusammenhang lautet, dass, wenn evaluative Begriffe semantisch an bestimmte Äußerungsformen geknüpft sind – wie ‚X ist gut‘ an ‚Ich empfehle X‘, oder ‚Ich soll X tun‘ an ‚Lass mich X tun‘ –, diese Äußerungen in rein syntaktischen Transformationen ihren Sinn ändern würden. So ist der Satz ‚X ist gut‘ klarerweise *keine* Empfehlung, wenn es als Antezedens eines Konditionals verwendet wird – wie z.B. in ‚Wenn X gut ist, dann ist auch Y gut‘. Dies hat Auswirkungen auf einfache Folgerungen wie z.B. die folgende:

- (P1) Wenn X gut ist, dann ist Y gut
- (P2) X ist gut.
- (K) Also ist Y gut.

Prämisse 2 würde in diesem Falle einfach nicht dasselbe *bedeuten* wie das Antezedens in Prämisse 1, da in 2 die Sprecherin nach nonkognitivistischer Interpretation eine Empfehlung vollzieht, in 1 jedoch überhaupt nichts empfohlen wird.<sup>259</sup> Wenn die Bedeutung jedoch verschieden wäre, wäre das Argument ungültig – es würde nämlich auf einer Äquivokation basieren –, was wiederum gegen die nonkognitivistische Interpretation der evaluativen Äußerung sprechen würde, da wir

---

<sup>257</sup> Vgl. die analoge Äußerung von R. Jay Wallace hinsichtlich der Verbindung von evaluativen und intentiven illokutionären Absichten. „This intuitive disanalogy between the practical and theoretical cases is reflected in the fact that first-person utterances of the following statements do not appear to be paradoxical at all: "I really ought to do *x*, but I'm going to do *y* instead," "X would be the best thing to do under the circumstances, but I intend to do *y*," "I've chosen to do *y*, though it's not in fact the best alternative open to me," etc." (Wallace 2001, 11f.) Vgl. die Diskussion bei Cholbi 2009.

<sup>258</sup> Vgl. Geach 1965; Frege 1993; Searle 1962; Searle 1969, 136–141; Dunn 1987, 30–39. Eine generelle Kritik an der sprechakttheoretischen Argumentation übt Hurka 1982.

<sup>259</sup> Auch bei reinen Behauptungen bezieht sich die behauptende Kraft auf das gesamte Konditional, nicht auf die einzelnen Glieder: Wer sagt ‚Wenn es regnet, dann ist der Boden nass‘ behauptet weder, dass es regnet, noch, dass der Boden nass ist, sondern nur, dass der konditionale Zusammenhang zwischen den beiden Umständen besteht. Natürlich ist die logische Analyse von Bedingungssätzen bis heute umstritten.

das Argument-Schema als gültig ansehen. Das Argument von Geach kann folgendermaßen verstanden werden:

- (P1) Logische Prädikate wie Konjunktion, Subjunktion, Implikation und Negation sind wahrheitsfunktionale Operatoren, d.h. sie bilden Wahrheitswerte auf andere Wahrheitswerte ab.
- (P2) Wenn logische Operatoren wahrheitsfunktional sind, so sind deren Argumente (Gegenstände) auch wahrheitsfunktional oder selbst wahr bzw. falsch.
- (P3) Einige evaluative Äußerungen sind Argumente (Gegenstände) von logischen Prädikaten.
- (K) Also: Einige evaluative Äußerungen sind wahrheitsfunktional oder selbst wahr bzw. falsch.

Nicht alle evaluativen Äußerungen benutzen logische Prädikate, daher ist die Konklusion nur eine Existenzbehauptung. Da diese Prädikate jedoch so elementar und einfach sind, betrifft dies faktisch eine Unzahl an evaluativen Äußerungen, und ist daher ein schwerwiegender Einwand gegen eine nonkognitivistische Semantik evaluativer Ausdrücke. Hare differenziert seinen Ansatz aufgrund solcher Argumente, die auf eine mögliche Multiplizierung von Bedeutungen evaluativer Terme im Nonkognitivismus hinweisen, indem er zwischen drei Bedeutungskomponenten unterscheidet: die bereits bekannte phrastische (phrastic) Komponente, die neustische (neustic), die nun nur mehr als Zeichen für die Äußerungsform (also deklarativ, direktiv, kommissiv etc.) steht, und die tropische (tropic) als Zeichen für die konventionell mit dem Zeichen verbundene Äußerungsform.<sup>260</sup> Der Sinn des Wortes ‚versprechen‘ würde demnach zu einem Teil darin bestehen, normalerweise einen Satz, in dem es verwendet wird, mit einer kommissiven Illokution (etwas versprechen) zu verbinden. Aber dies heißt nicht, dass das Wort ‚versprechen‘ in jedem Kontext dafür sorgt, dass ein Sprecher ein Versprechen gibt, wenn er es äußert, wie man z.B. anhand der Frage ‚Habe ich dir versprochen, dich von der Schule abzuholen?‘ leicht sehen kann. In dem obigen Konditional der Prämisse 1 würde demnach ‚X ist gut‘ im Antezedens seine neustische Komponente verlieren, seine tropische – *normalerweise* ist ‚gut‘ ein Zeichen für eine Empfehlung seitens des Sprechers – und seine phrastische Komponente (unter andern ‚X‘) bleiben aber dieselben. Auf diese Weise lässt sich das Frege-Geach Problem zumindest entschärfen, ob es gänzlich dadurch gelöst wird, kann ich hier nicht weiter verfolgen.<sup>261</sup>

---

<sup>260</sup> Vgl. Hare 1970, 20f.; Dunn 1987, 38ff.

<sup>261</sup> Für einen neueren Vorschlag zur Lösung des Frege-Geach-Problems im Rahmen einer expressivistischen Ethik vgl. auch Gibbard 1990; Blackburn 1984. Eine eingehende Diskussion des semantischen Programms des Expressionismus bietet Mark Schroeder 2008.

2) Imperative sind nicht dazu geeignet, um volitives Denken adäquat auszudrücken.<sup>262</sup> Hier wendet sich Dunn gegen imperative Theorien des Willens (imperative theory of the will), die Imperative als geeigneten Ausdruck für Intentionen, etwas zu tun, ansehen. Dagegen wendet Dunn ein, dass Imperative nur dafür geeignet sind, Intentionen auszudrücken, die das Handeln anderer Personen betreffen, nicht jedoch das eigene.<sup>263</sup> In den Fällen, in denen jemand einen Imperativ an sich selbst richtet, distanzieren sich die Person imaginativ von sich selbst, um durch einen Befehl an dieses so distanzierte Selbst sich zu einer Handlung zu bewegen.<sup>264</sup> Der Imperativ an sich selbst stellt demnach einen Umweg gegenüber dem einfachen Weg, die Handlung zu intendieren, dar, aber offensichtlich bereitet letzteres der Person Schwierigkeiten – sie kann sich z.B. nicht dazu ‚durchringen‘, etwas zu tun –, weshalb sie den Umweg über den Imperativ wählt.

3) Evaluative Ausdrücke haben generell kein volitives Bedeutungselement, sondern sind einfach Indikative.<sup>265</sup> Nach dieser erweiterten These sind evaluative Ausdrücke wie ‚Ich soll X tun‘ *semantisch* nicht mit expressiven (Ausdruck von Wünschen, etwas zu tun), kommissiven (als Ausdruck von Intentionen, etwas zu tun) oder direktiven (als Ausdruck von Wünschen oder Intentionen, dass andere etwas tun) Illokutionen verbunden. In der von Hare präferierten Analyse von ‚Ich soll X tun‘ als ‚Lass mich X tun‘ fungiert letzteres nicht als Ausdruck eines direktiven, sondern kommissiven (intentionalen) Sprechaktes.<sup>266</sup> Evaluative Terme *können* mit kommissiven Illokutionen, wie z.B. die Äußerung einer Drohung durch ‚Das solltest du lieber nicht tun!‘, verbunden sein, ebenso mit konativen Illokutionen, wie z.B. dem Ausdruck eines Wunsches durch ‚Heute wäre ein guter Tag zum Skifahren‘. Diese Phänomene sprechen aber nicht für eine notwendige qua semantische Verbindung solcher Illokutionen mit evaluativen Termen:

---

<sup>262</sup> Vgl. Dunn 1987, ch. 3.

<sup>263</sup> „Now it is of course true that directive illocutions, the illocutions with which imperatives are semantically connected, count as the expression of certain intentions or desires, and so independent imperatives are correctly identified as appropriate sentences in which to express such volitions. But what Hare ignores is that directive illocutions are essentially illocutions in which one person says to someone *else* that he should do something and therefore count as expressing intentions or desires on the part of speakers that *others* should act in certain ways, given that it is a sincerity condition on such illocutions that such other-person intentions or desires are present in speakers [...]” (Dunn 1987, 60f.)

<sup>264</sup> „Typical cases of the phenomena that go by the name of self-addressed directives reflect this fundamental fact about *bona fide* directive illocutions [dass sie wesentlich auf andere gerichtet sind]. For they, I suggest, are imitative of the latter as self-dissociative exercises of the imagination in which individuals imagine that they are issuing directives to someone *else*, in order to get themselves to act in certain ways.” (Dunn 1987, 66)

<sup>265</sup> Vgl. Dunn 1987, ch. 3. „Evaluatives apt for expressing all-out present-tense summary evaluative thinking about action are only what they undeniably are, namely, indicatives. They do not include, in addition, a relevant volitive semantical component. Therefore, the evaluative thinking such evaluatives are suitable for expressing is revealed, in the language of its expression, as wholly cognitive or theoretical and nonvolitional.” (Dunn 1987, 80f.)

<sup>266</sup> Vgl. Dunn 1987, 72ff.

„What I'm insisting on is that this kind of phenomenon requires an interpretation that does not mistakenly impute moods of the will to the evaluatives involved. For such illocutionary associations, I contend, are not in any way *semantically* guaranteed: speakers perfectly well can dissociate their evaluative illocutions from *any or all* of the relevant illocutionary forces.” (Dunn 1987, 78)

Dies zeigen wiederum Fälle, in denen Personen evaluative Terme gebrauchen und sich gleichzeitig von jeglichen intentiven oder konativen Illokutionen distanzieren können, ohne sich dabei selbst zu widersprechen oder Sinnloses zu äußern. Weiters lassen sich evaluative Terme auf Verben anwenden, die keine Handlungsverben sind, und deshalb nicht befohlen oder intendiert werden können, wie z.B. ‚Du solltest glauben, dass P‘ oder ‚Du solltest dich schämen!‘. Nach nonkognitivistischer Interpretation hat ‚soll‘ in diesen Sätzen eine andere Bedeutung als in anderen Sätzen, und es ist nicht verständlich, warum dies so sein sollte. Insgesamt sprechen somit einige schwerwiegende Gründe gegen die präskriptivistische Theorie von Willensschwäche.

### **III.4 Willensschwäche, Inkommensurabilität und Inkomparabilität**

Willensschwäche motiviert nicht nur Teilungstheorien des Geistes, sondern auch Teilungstheorien im Bereich der Werte. Werteinkommensurabilität soll verständlich machen, wie der Wert, der durch die willensschwache Handlung realisiert wird, nicht restlos in dem Wert der willensstarken Alternative ‚aufgeht‘ und so als Grund für diese Handlung dienen kann. Zuerst diskutiere ich David Wiggins Konzept, danach eine entsprechende Kritik von Michael Stocker.

#### **III.4.1 David Wiggins: Inkommensurable Werte und Willensschwäche**

David Wiggins argumentiert in seinem Aufsatz *Weakness of will, commensurability, and the objects of deliberation and desire* (1978) dafür, dass das Phänomen der Willensschwäche eng mit der Inkommensurabilität von Werten und Gütern zusammenhänge.<sup>267</sup> Die Inkommensurabilität erläutere, weshalb die willensschwache Handlung auch dann noch aus Sicht der Person attraktiv sei, wenn sie als schlechter gegenüber einer anderen Handlung beurteilt werde. Dies mache weiters verständlich, weshalb die willensschwache Handlung aus einem Grund vollzogen werde und damit nicht völlig grundlos sei. Wiggins beschreibt das Phänomen der Willensschwäche daher als Handlungen, die irrational, jedoch nicht völlig grundlos vollzogen werden:

„Almost anyone not under the influence of theory will say that, when a person is weak-willed, he intentionally chooses that which he knows or believes to be the worse course of action when he could choose the better course; and that, in acting in this way, the weak-willed man acts not for *no* reason at all – that would be strange and atypical – but irrationally.“ (Wiggins 1978, 251)

---

<sup>267</sup> Ähnlich argumentieren Burnyeat 1980; Nussbaum 1984; Nussbaum 1986, ch. 4, 89–121; Nussbaum 2001; Raz 1997.

Seine Argumentation knüpft an den antiken Diskussionen über Akrasia in Platons *Protagoras* und in der *Nikomachischen Ethik* des Aristoteles an.<sup>268</sup> Das sokratische Argument gegen die Existenz von Willensschwäche versteht er dabei folgendermaßen: Für den sokratischen Hedonisten existiert ein universeller Maßstab, anhand dessen alle anderen Wertqualitäten eindeutig gemessen werden können.<sup>269</sup> Das Maß ist die Menge an Lust, die ein Gegenstand spendet, und der dazugehörige Maßstab ist kardinal, d.h. er erlaubt neben einer ordinalen Rangordnung auch die Bestimmung der Abstände zwischen den gemessenen Gegenständen.<sup>270</sup> Ordinale Maßstäbe oder Skalen bilden lediglich eindeutige Ordnungen anhand eines Merkmales ab, wie z.B. die Rangfolge in einem Wettbewerb. Die Elemente werden dabei durch die Relationen ‚Identität‘ (oder ‚gleich gut‘ etc.), ‚größer als (oder ‚besser als‘ bzw. ‚stärker als‘ etc.) und ‚kleiner als‘ (oder ‚schlechter als‘ bzw. ‚schwächer als‘ etc.) gereiht. Kardinale Skalen messen zudem die Abstände zwischen den einzelnen Elementen, wie z.B. eine Datumsangabe oder Temperaturangaben. Wichtig ist, dass hier additive Operationen sinnvoll angewendet werden können. Auf einer Kardinalskala mit absolutem Nullpunkt (Verhältnisskala) lassen sich zudem auch Multiplikationen sinnvoll anwenden, wie z.B. beim Längenmaß. Die Kommensurabilitätsthese von Werten kann definiert werden als Existenzthese einer kardinalen Skala zwischen verschiedenen Werten. *Starke Inkommensurabilität* wäre die These, dass es zu *keinem* beliebigen Wertepaar *zumindest eine* kardinale Skala gibt, anhand derer beide messbar sind, *schwache Inkommensurabilität* hingegen verneint nur, dass es *eine* kardinale Skala gibt, die für *alle* beliebigen Werte anwendbar ist. Allgemein gilt, dass das Fehlen einer kardinalen Skala nicht das Fehlen einer ordinalen Skala impliziert, bzw. impliziert Inkommensurabilität nicht zugleich Inkomparabilität.<sup>271</sup> Kombiniert man die Existenz eines

---

<sup>268</sup> Die Adäquatheit der Interpretationen der beiden Texte werde ich hier nicht weiter hinterfragen. Vgl. Kap II.1; Kap. III.2.1.

<sup>269</sup> Vgl. Nussbaum 1984; Nussbaum 2001. Nussbaum veranschaulicht das Paradox willensschwachen Handelns aus Sicht eines sokratischen Hedonisten anhand folgenden Beispiels: „Phaedra’s rational principle, for some reason, is to maximize her bagel-eating. Standing in the middle of the room, she sees on a table on one side a plate containing two fresh bagels, toasted and buttered. On the other side of the room, on a similar table, is a plate containing one toasted buttered bagel. The bagels are the same variety, equally fresh, equally hot, buttered in the same way. She can go for either one plate or the other, but (for some reason) not both. She knows that, given her rational principle, she ought to eat the two bagels. But, overcome by desire, she eats the one. Now this does seem highly peculiar, in a way that our first case did not. We must comprehend that there is *no* respect in which the single-bagel plate differs from the two-bagel plate, except in the number of bagels it contains. The bagels are in no way qualitatively different. Nor is the arrangement of bagels on plate or plate on table somehow more aesthetically appealing. The single-bagel plate is not even nearer, it is the same distance away. What could make Phaedra’s choice anything but absurd, given that she really has the principle we say she has?“ (Nussbaum 1984, 62f.) Nussbaum nimmt hier an, dass unter Voraussetzung der Kommensurabilität aller Werte alle Faktoren der beiden Optionen, die motivational relevant sein können, auch evaluativ relevant sein müssen.

<sup>270</sup> Vgl. Chang 2001; Chang 1997.

<sup>271</sup> „Perhaps knowingly choosing the less valuable good is coherent if the goods are incommensurable. For us to be able even to state this possibility, the incommensurable goods must be *comparable*. If we can make sense of the claim that a person acts akratically in regard to two goods, it must make sense to say of the two things that the one is better than the other because it has a certain amount of one good rather than a certain amount of the other. If the goods

universellen kardialen Maßstabes für alle Güter mit der Annahme, dass ein solcher Hedonist stets als Ziel hat, seine Lust zu maximieren, ergibt sich die Ablehnung von Willensschwäche: wenn jemand bloß Lust sucht, und feststellt, dass von zwei Handlungen die eine mehr, die andere weniger Lust verspricht, sei es völlig unverständlich, wie er dennoch absichtlich die Handlung vollziehen könne, die weniger Lust verspreche.<sup>272</sup> Im Umkehrschluss muss der Hedonist folglich im Unwissen über den wahren Lustwert sein, der mit den jeweiligen Handlungsoptionen verknüpft ist.

„Let  $\phi$  be the universal or all-purpose predicate of favourable assessment. A man will only be incontinent if he knows or believes the thing he doesn't do is the thing with most  $\phi$  to it. But if that is the alternative that has most  $\phi$  to it, and if nothing else besides  $\phi$ -ness counts positively for anything, there is nothing to commend any other course of action over the one that is most  $\phi$ . He could have had no reason, *however bad*, for choosing the other. The choice of a smaller amount of pleasure now against a larger amount of pleasure later is explicitly described as a form of ignorance in the supposedly single dimension  $\phi$ ; [...] If everything with any relevance to choice is comprehended in the question how  $\phi$  a given course of action is, and how  $\phi$  its competitors are, then no rational sense can be made of weakness of will. This is the *Protagoras* argument.“  
(Wiggins 1978, 267)

Unter Inkommensurabilität versteht Wiggins das Fehlen eines gemeinsamen kardinalen Maßstabes, mit dessen Hilfe sich verschiedene Werte auf einer einzigen Skala messen ließen. Nach Ruth Chang lassen sich folgende Bedeutungen von Inkommensurabilität von Werten – abgesehen von Inkommensurabilität in Wissenschaftstheorie und Epistemologie – unterscheiden:<sup>273</sup>

- 1) *Inkompatibilität* von Werten: eine Handlung bzw. Wertrealisierung schließt eine andere aus; 2) *Wertepluralismus*: es existieren verschiedene, voneinander distinkte Werte; 3) *Diskontinuität* bzw. lexikalische Priorität: eine Reihung von Werten, in denen höhere Werte automatisch niedrigere Werte ‚übertrumpfen‘, d.h. dass die noch so geringe Realisierung des höheren Wertes einer noch so großen Realisierung des geringeren Wertes strikt vorzuziehen ist. 4) *Nichtsubstituierbarkeit* von Werten: die Realisierung eines Wertes wiegt nicht den Verlust der Nichtrealisierung eines anderen Wertes auf; 5) *Inkommensurabilität*: das Fehlen eines gemeinsamen Maßstabes, anhand dessen Werte gemessen werden können; 6) *Inkomparabilität* von Werten: Zwei Werte sind inkomparabel,

---

were not comparable, then all we could say of someone who chose one rather than the other would be that the person had done simply that: chosen one rather than the other. We could not say that the person had chosen the better or the worse. Nor therefore, could we say that the person acted akratically. (Thus, we can see that commensurability entails comparability.)“ (Stocker 1984, 270) Problematisch ist an dieser Beschreibung, dass davon ausgegangen wird, dass ein reiner Vergleich impliziere, dass das eine Element eine bestimmte *Menge* des einen Gutes im Vergleich zum anderen hätte: ein rein ordinaler Vergleich macht aber gerade keine Angaben über *bestimmte* Mengen, sondern reiht lediglich im Sinne eines mehr oder weniger (bzw. besser oder schlechter).

<sup>272</sup> „If everything trades off against everything else at some predeterminable (however contextually variable) rate and the trading rate reflects everything that matters about it, what reason could anyone have (and what reason could there be for him or her) to think twice about whether to do the act that scored highest or to backslide from the demonstrably optimistic act (unless, of course, he or she wanted to go back on the calculation)? It will be harder than it ought to be to understand weakness of will as *having reasons* (of a sort) *of its own*.“ (Wiggins 1997, 267, FN 13)

<sup>273</sup> Vgl. Chang 2001.

wenn keine Ordnungsrelation zwischen ihnen besteht, d.h. der eine Wert weder besser als, noch schlechter als, noch gleich gut wie der andere Wert ist.<sup>274</sup>

Wiggins verknüpft die Inkommensurabilitätsthese mit der Ablehnung der These der Kompensation gleicher Art (compensation in kind), d.h. der Aufwiegung eines Werteverlustes durch einen gleichartigen und zumindest gleichhohen Wertegewinn: Wenn X besser ist als Y, dann gibt es keine Eigenschaft Z an Y, die nicht gleichwertig oder besser in X aufgewogen wird.<sup>275</sup> Der Verlust von 50 € wäre nach diesem Prinzip vollständig in dem Gewinn von 100 € aufgewogen, sowie die Menge an Lust, welche die eine Handlung verspricht, vollständig durch die Menge an Lust, welche die andere Handlung verspricht, aufgewogen wäre, wenn Lust ein universeller Maß für den Vergleich von verschiedenen Handlungen und deren Werte wäre. Da Werte jedoch nach Wiggins schwach inkommensurabel sind und sich das Prinzip des Kompensation gleicher Art daher nicht auf sie anwenden lässt, ist der Verzicht auf die willensschwache Handlung mit einem *spezifischen Verlust* verbunden, der nicht durch die bessere Alternative aufgewogen werden kann.

„But, so soon as we clearly formulate this principle [of compensation in kind] and see that it is false, I think we shall see room for the *akrates* to choose the smaller good when he could have the larger, and choose it for a reason which is a *real* reason for all that it is a *bad* reason. This can happen wherever there is no prospect of compensation in kind, and y has some peculiar or distinctive charm that the incontinent man is susceptible to.“ (Wiggins 1978, 269)

Die Argumentation von Wiggins kann demnach folgendermaßen als modus tollens rekonstruiert werden:

- (P1) Wenn alle Werte kommensurabel sind, dann lassen sich alle Verluste und Gewinne an Werterfüllungen zweier Wertträger gegeneinander eindeutig aufwiegen. (These der Kompensation gleicher Art bzw. der Substituierbarkeit von Werten – compensation in kind)
- (P2) Wenn sich alle Verluste und Gewinne an Werterfüllungen zweier Wertträger gegeneinander eindeutig aufwiegen lassen, dann hat die willensschwache Person keinen Grund, willensschwach zu handeln
- (P3) Die willensschwache Person hat aber einen Grund, willensschwach zu handeln
- (K) Also: Nicht alle Werte sind kommensurabel bzw. einige Werte sind inkommensurabel (nicht kommensurabel)

---

<sup>274</sup> Zur näheren Diskussion über Inkommensurabilität vgl. die Einleitung zu Chang 1997.

<sup>275</sup> Die implizite Definition der These von Wiggins anhand seiner Beschreibung der eudaimonia bei Aristoteles: „Nor does it imply that if course x is better in respect of *eudaimonia* than course y, then there is no desirable feature that y offers that x does not offer too, by way of an equal or greater degree of that very feature. One might call this the principle of compensation in kind.“ (Wiggins 1978, 269) Wiggins vertritt eine inklusive Interpretation der eudaimonia statt einer dominanten, also die Ansicht, eudaimonia werde durch verschiedene intrinsische Güter konstituiert, anstatt als dominantes Ziel allen menschlichen Strebens zu gelten, für die andere Güter lediglich instrumentellen Wert besitzen. Vgl. Stocker 1984, 271; Wiggins 1978, ch. VI.

Die These der Inkommensurabilität von Werten soll damit verständlich machen, wie a) die willensschwache Person aus einem Grund handelt und b) die *willensstarke* Person dennoch Bedauern über den Verlust der mit der unerfüllten Handlungsalternative verbundenen Wert-erfüllung empfinden kann.

An Prämisse P1 ist fraglich, ob die Kompensation von Verlusten bzw. die Substituierbarkeit von Gütern tatsächlich mit der Kommensurabilität von Werten einhergeht. Nur weil *Werte* kommensurabel sind, sind nicht notwendigerweise die mit ihnen verbundenen *Güter* auch zueinander kommensurabel. Außerdem kann ein abstrakter Wert durch mehrere Faktoren bestimmt werden, und die Kompensation des Wertes durch einen anderen muss damit nicht mit der Kompensation der einzelnen Faktoren einhergehen, die mit dem Wert verknüpft sind. So könnte Lust durch die Faktoren Intensität und Dauer bestimmt sein, wodurch das gleiche Ausmaß an Lust durch verschiedene Kombinationen der Faktoren Intensität und Dauer zustande kommen könnte. Beide Optionen hätten denselben Wert an Lust, aber motivational könnte wegen der verschiedenen Zusammensetzung der einzelnen Faktoren ein Unterschied zwischen den beiden Optionen bestehen. Allgemein ist außerdem die Frage nach der Kommensurabilität von Werten zu unterscheiden von der metaphysischen Frage nach der Pluralität von Werten. Jemand, der meint, es gebe im Grunde nur eine Art von Wert – z.B. ‚Gut-Sein‘ – muss nicht notwendigerweise die These vertreten, dass verschiedene Güter miteinander hinsichtlich dieses einen Wertes kommensurabel wären, da z.B. genaue Wertabstände nicht messbar sein könnten. Andererseits kann ein Wertpluralist meinen, dass alle Werte miteinander kommensurabel sind, ohne damit schon seinen Wertpluralismus aufzugeben. Diese Fragen hängen sehr stark davon ab, wie einzelne Werte überhaupt individuiert werden, und ich kann sie hier nicht weiter verfolgen.<sup>276</sup>

Prämisse P2 ist ebenfalls fragwürdig: Sie setzt implizit ein enges Verhältnis zwischen Werten und Gründen voraus, und des Weiteren ein enges Verhältnis zwischen Motivation und Werten bzw. zwischen affektiver Attraktivität, den mit dieser verbundenen Wünschen und Werten.<sup>277</sup> Die zugrundeliegende These ist, dass die Stärke der Motivation mit der Rangfolge der Wertigkeit korreliere. Weshalb sollte die willensschwache Person aber nur dann einen Grund für die willensschwache Handlung haben, wenn ihre Motivation, entsprechend zu handeln in Werten gründet, die mit der Handlung verbunden und inkommensurabel im Vergleich zu den Werten der alternativen Handlung sind? Und andererseits ist fraglich, weshalb die Attraktivität der willensschwachen Option schon dadurch verständlicher wird, indem sie mit Werten verbunden sind, die inkommensurabel sind mit denjenigen der willensstarken Option. Die Kritik wäre

---

<sup>276</sup> Vgl. Chang 2001.

<sup>277</sup> Vgl. Kap. IV.3.

demnach, dass die eigentliche Erklärung für Willensschwäche nicht die These der Inkommensurabilität leistet, sondern die These, dass die Attraktivität bzw. Motivation für Handlungsoptionen nicht eindeutig korreliert ist zu ihrer Evaluation: Ohne den spezifischen Charme („peculiar or distinctive charm“), der mit der willensschwachen Handlung verknüpft ist, wäre nicht verständlich, weshalb die Person motiviert ist, sie auszuführen. Wiggins verknüpft daher mit der These der Inkommensurabilität von Werten die These der *Fragmentierung der psychischen Ressourcen* der Motivation:

„Incommensurability was introduced in conjunction with the idea of the falsity of the principle of strict compensation in kind, and in order to suggest the heterogeneity of the psychic sources of desire satisfaction and evaluation (both the evaluation consequential upon desire and the evaluation which desire is consequential upon).“ (Wiggins 1980, 262)

Die reine Inkommensurabilitätsthese erläutert demnach noch nicht die Möglichkeit von Willensschwäche. Denn genauso wie der Verlust der Werterfüllung der willensschwachen Handlung nicht aufgewogen werden kann durch den Gewinn, der mit der besseren Handlung verbunden ist, kann auch der Verlust der Werterfüllung der besseren Handlung nicht aufgewogen werden durch den Gewinn, der mit der willensschwachen Handlung verbunden ist. Wenn beide Werte inkommensurabel zueinander sind, dann stellen *beide* Handlungsoptionen einen spezifischen Gewinn und Verlust dar, der nicht durch die jeweils andere Option aufgewogen werden kann. Was hinzukommt ist der besondere Reiz bzw. die besondere Attraktivität, die mit der willensschwachen Handlung verbunden ist, die nicht mit der gesamten Bewertung der beiden Handlungsoptionen korreliert. Folgendes Problem stellt sich in diesem Zusammenhang:<sup>278</sup> Willensschwäche setzt voraus, dass die Person die beiden alternativen Handlungen zumindest miteinander vergleichen kann im Sinne eines Urteils über die Relation ‚ist besser als‘, die zwischen den beiden Optionen besteht. Trotz der Inkommensurabilität der Werte sind die beiden Optionen demnach komparabel. Jeder Vergleich macht nur Sinn im Hinblick auf etwas, hinsichtlich dessen es verglichen wird. Im besten Falle besteht der Vergleich hinsichtlich aller Gründe, die eine Person zur Verfügung hat: Die eine Handlung x ist *alles in allem gesehen* besser als

---

<sup>278</sup> „The incommensurable values are either incomparable or comparable. If they are incomparable, their acts are not within the ambit of maximization: neither can be said to be better than the other, nor can either be taken as the better or best act. But since neither is lesser than the other, it is not akratic to do one instead of the other. Suppose now that the incommensurable values are comparable. Now one of the acts can be taken as better and one as lesser. But given psychological maximization, we are faced with exactly the same problems in seeing how coherent akrasia is possible as we did when there was only one value. For according to psychological maximization, the better act, even though its values are incommensurable with those of the lesser act, must be desired more than the lesser act. And on the assumption that we do what we most want to do – or that, especially because we are concerned with akratic acts, we are now concerned with acts done from the greater desire – the lesser act cannot be done. At the least, it cannot be done from a greater desire. To summarize: if because of incomparability, acts with incommensurable values cannot be ranked, no act can be taken as lesser nor can any be akratic. But if they can be ranked, then on the assumption of psychological maximization, the lesser cannot be done, nor therefore is akrasia possible. Or rather, the lesser cannot be done, nor can an akratic act be done, from a greater desire.“ (Stocker 1984, 236)

die alternative Handlung  $y$ . Es braucht demnach eine Erläuterung dafür, weshalb sich die Motivation der willensschwachen Person für die Handlung, die schlechter bewertet wurde, nicht gemäß der evaluativen Bewertung der beiden Optionen verhält, d.h. noch immer attraktiv ist. Die reine Inkommensurabilitätsthese kann dies nicht verständlich machen *ohne* die These der Fragmentierung der psychischen Ressourcen und ist damit nicht hinreichend dafür, die Möglichkeit von Willensschwäche zu erläutern. Michael Stocker argumentiert außerdem dafür, dass bereits die Fragmentierung der psychischen Ressourcen alleine hinreichend ist, um Willensschwäche möglich zu machen, und die Inkommensurabilitätsthese daher dafür auch nicht notwendig sei.

### III.4.2 Michael Stocker: Willensschwäche trotz Kommensurabilität

Trotz seiner generellen Argumentation für einen Wertepluralismus in seinem Buch *Plural and conflicting values* (1990) kritisiert Michael Stocker die Argumentation von David Wiggins, die These der Inkommensurabilität sei notwendig, um die Möglichkeit von Willensschwäche zu erläutern.<sup>279</sup>

„Many philosophers, including Wiggins, hold that coherent akrasia – i.e. akrasia which escapes the Protagorean Predicament<sup>280</sup> – requires incommensurable plural values. In this chapter, I argue that this claim is mistaken and that it requires virtually unusable notions of unity, plurality, and commensurability; that it trades on unacceptable understandings of action and reason; and that we can have coherent akrasia in regard to one value.“ (Stocker 1990, 212)

Willensschwäche in Bezug auf einen einzigen Wert könne dann kohärent gedacht werden, wenn zwischen einer *affektiven* und einer *kognitiven* Bezugnahme unterschieden wird. So könnte selbst ein vollkommener Hedonist eine Handlungsoption attraktiver finden, die er zugleich für weniger wertvoll, da weniger lustbringend, einschätzt. Dies sei dann möglich, wenn Lust aus komplexen Momenten wie zeitliche Dauer, Intensität, Schärfe usw. konstituiert sei, die verschiedene Attraktivität verständlich mache.

„Intensity is clearly not the only dimension of pleasure which is proper to pleasure. There are the dimensions discussed by Bentham – including that most quantitative one, quantity. As well, there are all the more purely phenomenological modalities – e.g. piquancy, sharpness, languorousness. All of these allow for a distinction between pleasures. They allow for different objects of attraction, charm, fascination. And more importantly, they allow for the degree of such affectively felt attraction, charm, and fascination to diverge from the degree of cognitively seen pleasure. It is possible to be so charmed by the particular piquancy, say, of an act's pleasure that one does that act even though another and even more pleasurable act is seen to be possible.“ (Stocker 1984, 221)

Wenn aber, so das Argument, selbst ein Hedonist bezogen auf einen einzigen Wert willensschwach sein kann, da seine affektive und kognitive Bezugnahme auf dasselbe Objekt

---

<sup>279</sup> Vgl. Stocker 1984; Stocker 1990; Pugmire 1982, 183ff.

<sup>280</sup> Das protagoreische Dilemma besteht aus dem scheinbar inkohärenten Konzept eines Handelnden, der ein Ziel verfolgt und zudem glaubt, auf eine bestimmte Weise dieses Ziel besser erreichen zu können als durch andere Handlungen und dennoch nicht entsprechend handelt. Vgl. Stocker 1990, 211f.

einander nicht notwendigerweise decken, so braucht es die These der Inkommensurabilität von Werten nicht, um Willensschwäche verständlich zu machen.<sup>281</sup> Eine hedonistische Theorie könnte, um diese Folgerung zu vermeiden, annehmen, dass nur Lust attraktiv sei, oder sogar nur maximale Lust.<sup>282</sup> Der Vorwurf wäre dann, dass der Quasi-Hedonist von zuvor, für den nicht nur Lust als Lust, sondern auch die besondere Schärfe einer Lust attraktiv sein kann, eigentlich nur die These bestätige, dass Inkommensurabilität von Werten notwendig sei für die Erläuterung willensschwachen Handelns, da sie Voraussetzung sei für die Unterscheidung verschiedener Arten von Lust.<sup>283</sup> Dagegen argumentiert Stocker, dass sein Argument nur voraussetze, dass sich die Objekte der Anziehung für den Akratiker in einem nicht völlig trivialem Sinne voneinander unterscheiden – die willensschwache Alternative ist verschieden von der als besser erachteten Wahlmöglichkeit. Wer nun argumentiert, dass diese Unterscheidung der Objekte bereits eine Pluralität und Inkommensurabilität von Werten voraussetze, der argumentiert eigentlich nicht für den Spezialfall der Willensschwäche, sondern für jede Art von Entscheidungssituation zwischen Objekten, die in irgendeiner nicht-trivialen Weise voneinander unterschieden wären.<sup>284</sup> Dies ist noch keine triftiger Grund gegen diese These, aber sie erweitert den Anspruch und die Reichweite der ursprünglichen Argumentation beträchtlich: statt den Spezialfall der Willensschwäche zu lösen, sei Inkommensurabilität nun notwendig zur Erläuterung der Möglichkeit jeglicher Entscheidung.

Stocker führt noch andere Gründe an, um seine Annahme, dass Attraktivität und Evaluation nicht notwendigerweise miteinander korreliert seien, zu stützen. Bereits in *Desiring the bad* (1979) hat Stocker dafür argumentiert, dass Evaluation und Motivation in einem komplexen statt einem einfachen, korrelativen Verhältnis zueinander stünden:

„I shall argue that motivation and evaluation do not stand in a simple and direct relation to each other, as so often supposed. Rather, they are interrelated in various and complex ways, and their interrelations are mediated by large arrays of complex psychic structures, such as mood, energy, and interest. Philosophical

---

<sup>281</sup> Vgl. Stocker 1984, 218ff.

<sup>282</sup> Vgl. Stocker 1984, 222ff.

<sup>283</sup> „These points made, we can put the objection against my use of the variegated hedonism this way: my argument claims that even with a field of commensurable values such as pleasure there can be coherent akrasia; but it works, or seems to work, only because it, albeit covertly, relies on incommensurability – e.g. between languorous pleasures as such and piquant pleasures as such. Thus it would be held that my argument sustains, rather than confutes, the claim that the distinction between commensurability and comparability is important for coherent akrasia. However, this worry is importantly self-refuting. For it holds only to the extent that the notions of unity and commensurability it involves are unimportant for akrasia – and indeed, for much else. My claim remains that the distinction between commensurability and comparability is unimportant for akrasia.” (Stocker 1984, 224)

<sup>284</sup> „The argument is thus misleading in two important and related ways. First, it is offered as an argument about the object of akratic choice. But it is really about something far more general: the object of non-degenerate differential choice. Second, it is offered as an argument that an evaluative monism is inadequate for akrasia. But, if right, it would show something far more general: that an evaluative monism is inadequate for non-degenerate differential choice.” (Stocker 1984, 229)

theories have ignored or misunderstood these structures and the corresponding all too common psychological phenomena. They have depicted the psyche, especially the interrelations between motivation and evaluation, as far too simple, far too unified, and far too rational." (Stocker, 1979, 738f.)

Er argumentiert dabei konkret gegen die Thesen, dass 1) das Gute notwendig attraktiv sei und 2) *nur* das Gute attraktiv sei, das Schlechte als solches also niemals Objekt eines Strebens sein könne. Die philosophische Debatte zu beiden Thesen ist entsprechend umfangreich und kann hier nicht annähernd adäquat diskutiert werden.

Gegen die erste These sprechen nach Stocker Fälle von Personen, die trotz intakter ‚Sicht‘ auf die evaluative Dimension bestimmter Objekte oder Handlungen diese nicht mehr oder zu wenig erstreben. Ob solche Personen tatsächlich vorstellbar sind, ist strittig und einer der Angelpunkte der Diskussion.

„Lack of this desire is commonplace. Through spiritual or physical tiredness, through accidie, through weakness of body, through illness, through general apathy, through despair, through inability to concentrate, through a feeling of uselessness or futility, and so on, one may feel less and less motivated to seek what is good. One's lessened desire need not signal, much less be the product of, the fact that, or one's belief that, there is less good to be obtained or produced, as in the case of a universal Weltschmerz. Indeed, a frequent added defect of being in such "depressions" is that one sees all the good to be won or saved and one lacks the will, interest, desire, or strength." (Stocker 1979, 744)

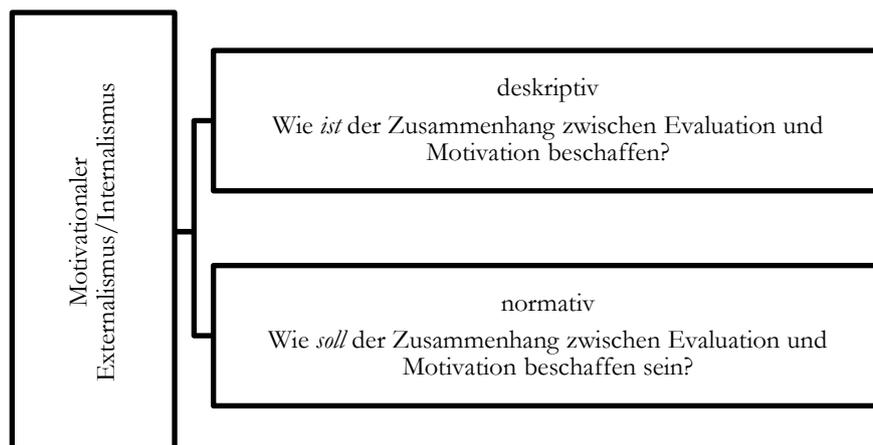
Gegen die zweite These sprechen nach Stocker Emotionen wie Hass und Eifersucht, die als direktes Objekt der auf sie folgenden Strebung die Verletzung oder Zerstörung einer Person oder eines Gegenstandes haben. Anstatt diese Strebungen so umzudeuten, dass sie in das Schema passen, das einzig das Gute attraktiv sei und erstrebt werden könne, sei es einfacher, als Objekt dieser Strebungen etwas aufzufassen, dass die betreffende Person selbst als schlecht beurteilt. Dasselbe gelte auch für selbstzerstörerische Wünsche, die gegen das eigene Wohl gerichtet sind.

„Given such moods and circumstances, harming another can be the proper and direct object of attraction. There is no need to posit another object, especially not an egoistic object like pleasure, power over others, showing oneself powerful, getting things to go one's own way, getting revenge. Just as there are desires and appetites directed at harming others, there are desires and appetites directed at harming oneself. In certain moods, such as the self-directed modes of disgust, hatred, guilt, shame, I may seek to humble, abase, or harm myself." (Stocker 1979, 748)

Die Vorstellbarkeit und Plausibilität der Erläuterung solcher Emotionen und Strebungen ist strittig. Beide Einwände zeigen jedoch, dass die beiden klassische Annahme, alles Streben erfolge *sub specie boni*, d.h. richte sich auf etwas aus der Perspektive der Person Gutes, und nur etwas Gutes könne erstrebt werden, vernünftigerweise bestritten werden können.

### III.5 Motivationaler Internalismus/Externalismus und Willensschwäche

Auf einer abstrakten Ebene wird das Verhältnis zwischen Evaluation und Motivation in der Debatte um den motivationalen Internalismus bzw. dessen Widerpart, den motivationalen Externalismus, diskutiert.<sup>285</sup> Es lassen sich allgemein zwei Formen unterscheiden: Externalismus und Internalismus bezogen auf Urteile (*Urteils-Internalismus/Externalismus*) und Externalismus und Internalismus bezogen auf Gründe (*Gründe-Internalismus/Externalismus*).<sup>286</sup> Theorien der ersten Form betreffen den Zusammenhang zwischen bestimmten Arten von Urteilen und motivationalen Zuständen von Personen. Theorien zweiter Form sind Theorien darüber, was gute Gründe, eine Handlung zu tun, eigentlich sind und in welchem Verhältnis diese zu motivationalen Zuständen von Personen stehen. Zusätzlich können die Theorien den jeweiligen Zusammenhang *normativ* oder *deskriptiv* verstehen, also als Aussage darüber, wie das Verhältnis von Urteilen oder Gründen und der Motivation zu handeln tatsächlich beschaffen *ist*, oder wie es beschaffen sein *sollte*.



Dieser Unterschied wird beizeiten durch die Hinzufügung von weiteren Bedingungen zu rein deskriptiven Theorien unterlaufen, die so normativ aufgeladen werden. So ist unklar, ob eine Theorie, welche beschreibt, wie bei *rein rationalen* Wesen der Zusammenhang zwischen Evaluation und Motivation beschaffen ist, für *nicht rein rationale* Wesen normativ zu verstehen ist oder nicht. Mit normativen Varianten von Internalismen werde ich mich näher in Kap. VI.1 beschäftigen. Folgend diskutiere ich nur das Verhältnis von deskriptiven Formen von Urteils-Internalismen zum Phänomen der Willensschwäche. Diese Theorien sind zudem eingebettet in verschiedene

<sup>285</sup> Diese Debatte ist zu unterscheiden von ähnlichen Debatten um internalistische und externalistische Positionen in der Epistemologie und Sprachphilosophie. Der Einfachheit wegen bezeichnen, falls nicht anders vermerkt, im Folgenden ‚Internalismus‘ und ‚Externalismus‘ immer die jeweiligen motivationstheoretischen Positionen.

<sup>286</sup> Bei Stephen Darwall ist dies die Unterscheidung zwischen ‚existence internalism‘ und ‚judgments internalism‘, vgl. Darwall 1983; Darwall 1992. Bei Derek Parfit ist dies die Unterscheidung zwischen ‚reason internalism‘ und ‚belief internalism‘, vgl. Parfit 1997. Bei David Brink findet sich die Unterscheidung zwischen ‚internalism about reasons‘ und ‚internalism about motives‘, vgl. Brink 1989. Vgl. auch Rosati 2008; Svavarsdottir 1999; Shafer-Landau 2003, part III; Björklund & al. 2012; Wedgwood 2007, ch. 1.

metaethische Diskussionen zwischen Kognitivismus und Nonkognitivismus, ethischen Realismus und Anti-Realismus, Deskriptivismus und Expressivismus u.a.m. Diese anderen Diskussionen sollen hier, soweit als möglich, ausgeblendet werden. Um den Rahmen der Diskussion sinnvoll zu beschränken, wird es in Folge auch lediglich um solche Urteile gehen, welche die eigene Person und das eigene Handeln betreffen. Zuerst unterscheide ich verschiedene Versionen von motivationalen Urteils-Internalismen (1), bevor ich für diese Positionen ein Dilemma konstruiere: in ihrer proportionalistischen Variante können Urteils-Internalismen synchrone motivationale Willensschwäche nicht in ihre Theorie integrieren, ohne trivial zu werden, wohingegen sie in ihrer nicht-proportionalistischen Variante ihrem eigenen Anspruch, den handlungsleitenden Charakter von evaluativen und normativen Urteilen zu erläutern, nicht gerecht werden. Nachdem ich kurz eine extreme Variante von Urteils-Internalismus von Sergio Tenenbaum diskutiere (3), stelle ich schließlich eine Version eines Urteils-Externalismus vor und erläutere näher die motivationalen Quellen des Handelns (4).

### III.5.1 Motivationaler Internalismus

Die generelle Idee hinter internalistischen Theorien von normativen oder evaluativen Urteilen<sup>287</sup> ist, die *handlungsleitende* Eigenschaft solcher Urteile plausibel zu machen: Wenn Personen darüber ein Urteil fällen, was sie tun sollen, dann sind sie normalerweise zu einem gewissen Grade motiviert, auch entsprechend zu handeln. Nach einem Internalismus besteht dieser Zusammenhang notwendig und er zeichnet praktische Urteile gegenüber bloß theoretischen Urteilen aus. Es gäbe demnach eine interne, da begrifflich notwendige Verbindung zwischen evaluativen Urteilen über Praktisches und der Motivation, entsprechend zu handeln.<sup>288</sup> Demgegenüber ist der Externalismus die These, dass keine derartige begrifflich notwendige Verbindung besteht. Aus argumentationsstrategischer Sicht sind die grundlegenden Schwierigkeiten von Internalismus und Externalismus tendenziell verschieden: Während internalistische Theorien Probleme mit Fällen von motivationalen Fehlleistungen haben, bereitet externalistischen Theorien die Erklärung des Normalfalles von geglückter Motivation Schwierigkeiten:

„Internalism and externalism face opposite challenges: internalists have an easy time with motivational success, but difficulty with motivational failures, whereas externalists make failure easily intelligible at the cost of struggling to explain success.“ (Kauppinen 2010, 236)

Bezogen auf Willensschwäche bedeutet dies, dass internalistische Theorien Probleme mit der Integration des Phänomens haben, während externalistische Theorien Schwierigkeiten damit haben, den problematischen Charakter von Willensschwäche adäquat zu erläutern. Ursprünglich

---

<sup>287</sup> Evaluative Urteile zielen auf Werte ab (X ist gut/schlecht/böse etc.) und normative Urteile auf Normen (Ich soll/darf/muss etc. X tun).

<sup>288</sup> Dieses Grundmuster findet sich in den zuvor besprochenen Theorien, vgl. Kap. III.1–4.

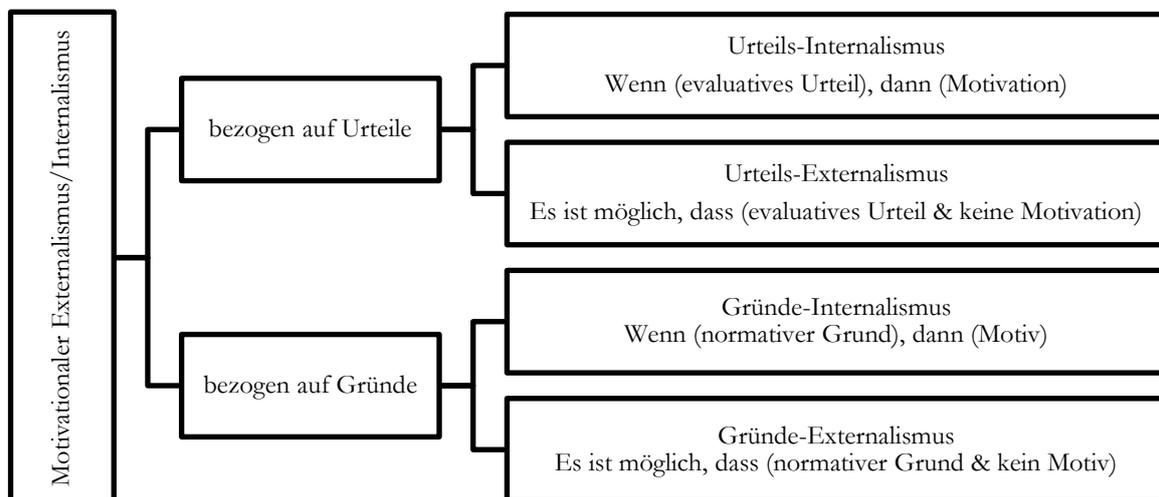
betrifft diese Diskussion vorwiegend moralische Urteile und deren Verknüpfung zu motivationalen Zuständen, wurde jedoch später auf weitere Arten von Urteilen ausgedehnt. Dies ist wichtig, da sonst Willensschwäche nur in Gestalt der moralischen Schwäche (*moral weakness*) für diese Diskussion als relevant betrachtet werden könnte. Eine mögliche Grundformulierung des Urteils-Internalismus lautet:<sup>289</sup>

Wenn eine Person aufrichtig urteilt/glaubt, dass sie F tun sollte, dann ist sie auch motiviert dazu, F zu tun.

In schematischer Form lässt sich die Urteils-Internalismus-These folgendermaßen darstellen:<sup>290</sup>

Wenn [evaluatives/normatives Urteil ( $\varphi$ )], dann [Motivation ( $\varphi$ )]

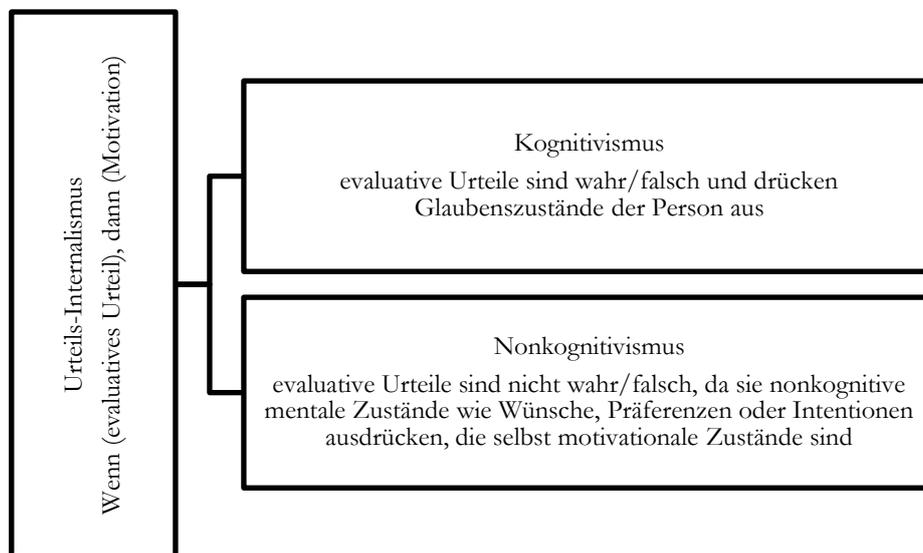
Dieses Schema kann nun verschieden ausgefüllt werden und dementsprechend ergeben sich verschiedene substantielle Versionen des Urteils-Internalismus. Dabei beantworten diese vor allem folgende drei Fragen: 1) Welche mentalen Akte und Zustände sind involviert in einem evaluativem Urteil? Sind es Überzeugungen und demnach *kognitive* Zustände oder *konative* und mithin nonkognitive Zustände? 2) Was an dem evaluativen Urteil ist genau begrifflich verknüpft mit der Motivation, entsprechend zu handeln? Ist es der *Gehalt* solcher Urteile oder der *Akt* des evaluativen Urteilens? 3) Welche Typen mentaler Zustände sind motivationale Zustände? Sind dies nur in einem weiten Sinne Pro-Einstellungen, oder können auch reine Glaubenszustände von sich aus zu Handlungen motivieren? Sind reine Überzeugungen nicht hinreichend und Pro-Einstellungen oder Wünsche notwendig für die Motivation zu handeln?



<sup>289</sup> Vergleiche entsprechende Formulierungen für moralische Urteils-Internalismen: „*Simple internalism*: Necessarily, if a person judges that she morally ought to  $\phi$ , then she is (at least somewhat) motivated to  $\phi$ .” (Björklund & al. 2012, 125) „Necessarily, if one is rational, then, if one judges ‘I ought to  $\varphi$ ’, one also intends to  $\varphi$ .” (Wedgwood 2007, 25) „If an agent judges that it is right for her to  $\phi$  in circumstances C, then she is motivated to  $\phi$  in C.” (Smith 1994, 61) „It is necessary and a priori that for any x and any y: if x judges that some available action y is morally right (good, obligatory, etc.), then x will have some (defeasible) motivation in favour of performing y.” (Joyce 2002, 337)

<sup>290</sup> In diesem Schema fehlen: ein modales Element (‘notwendigerweise’), eine Aufrichtigkeitsbedingungen (‘aufrichtig’) und Ausnahmeklauseln (‘oder praktisch irrational’ etc.), auf die ich später noch eingehe.

1) Evaluative Urteile lassen sich kognitiv oder nonkognitiv bzw. expressiv verstehen. Als kognitive Elemente kommen im weitesten Sinne glaubensartige Zustände in Frage, d.h. mentale Akte mit propositionalen Gehalten, die wahr oder falsch sein können. Diese können repräsentiert werden durch *Urteilsakte* bzw. die Zustimmung zu Urteilen (moralischen, prudentiellen oder anderen Gehalts), mithin mentale Episoden, oder durch Glaubenszustände (Dispositionen) bzw. Urteile<sup>291</sup> mit denselben Inhalten. Nonkognitive bzw. expressive Deutungen von evaluativen Urteilen verstehen diese als Ausdrücke von nicht glaubensartigen Zuständen oder zumindest nicht als rein kognitiv. Evaluative Urteile sind demnach weder wahr noch falsch, sondern drücken bestimmte Einstellungen von Personen zu Propositionen aus. Mit dem Urteil, eine Handlung sei gut, drücke eine Person lediglich ihren Wunsch, ihre Präferenz, ihre Zustimmung oder ihre Intention aus, dass andere Personen ebenso handeln bzw. dass sie selbst so handelt. Da nonkognitive Zustände diesem Verständnis nach bereits motivationale Zustände sind, ist der Urteils-Internalismus in dieser Lesart trivialerweise wahr.<sup>292</sup> Für Richard Hare z.B. sind moralische Urteile in ihrem präskriptiven Anteil einfach Ausdruck von Intentionen oder Imperativen, mithin bereits motivationale Elemente. Für das motivationale Element kommen alle Arten von Strebungen in Frage, also Zustände des Wünschens oder Absehens, die auf eigene Handlungen gerichtet sind.



<sup>291</sup> ‚Urteil‘ ist zweideutig zwischen *Produkt* von Urteilen und dem *Prozess* des Urteilens.

<sup>292</sup> Trivialerweise da die Einsetzung in das Schema ergibt: Wenn [Motivation ( $\varphi$ )], dann [Motivation ( $\varphi$ )]. Vgl. aber Joyce 2002, der dafür argumentiert, dass ein Expressivismus nicht einen Urteils-Internalismus impliziere. Wenn die Erfolgskriterien für das Ausdrücken eines Urteils, also ob jemand tatsächlich ein evaluatives Urteil gefällt hat oder nicht, linguistisch sind (ein Sprechakt wird ausgedrückt) und nicht mentalistisch (ein bestimmter mentaler Zustand liegt vor), dann könne eine Person aufrichtig ein evaluatives Urteil fällen und damit bestimmte motivationale Zustände ausdrücken ohne sie zu haben, mithin ohne motiviert zu sein, entsprechend zu handeln. Vgl. auch Joyce 2009; Ridge 2006. Dieser Punkt ist erst dann einleuchtend, wenn klar ist, in welchem Zusammenhang Sprechakte zu mentalen Akten stehen.

Nonkognitivisten haben den Urteils-Internalismus mitunter als Argument gegen kognitivistische Theorien verwendet. Schematisch lässt sich dieses Argument folgendermaßen verstehen:

- (P1) Evaluative Urteile motivieren notwendigerweise. (Urteils-Internalismus)
- (P2) Überzeugungen motivieren nicht notwendigerweise (Humeanische Theorie der Motivation).
- (K) Also: Evaluative Urteile drücken keine Überzeugungen aus.<sup>293</sup>

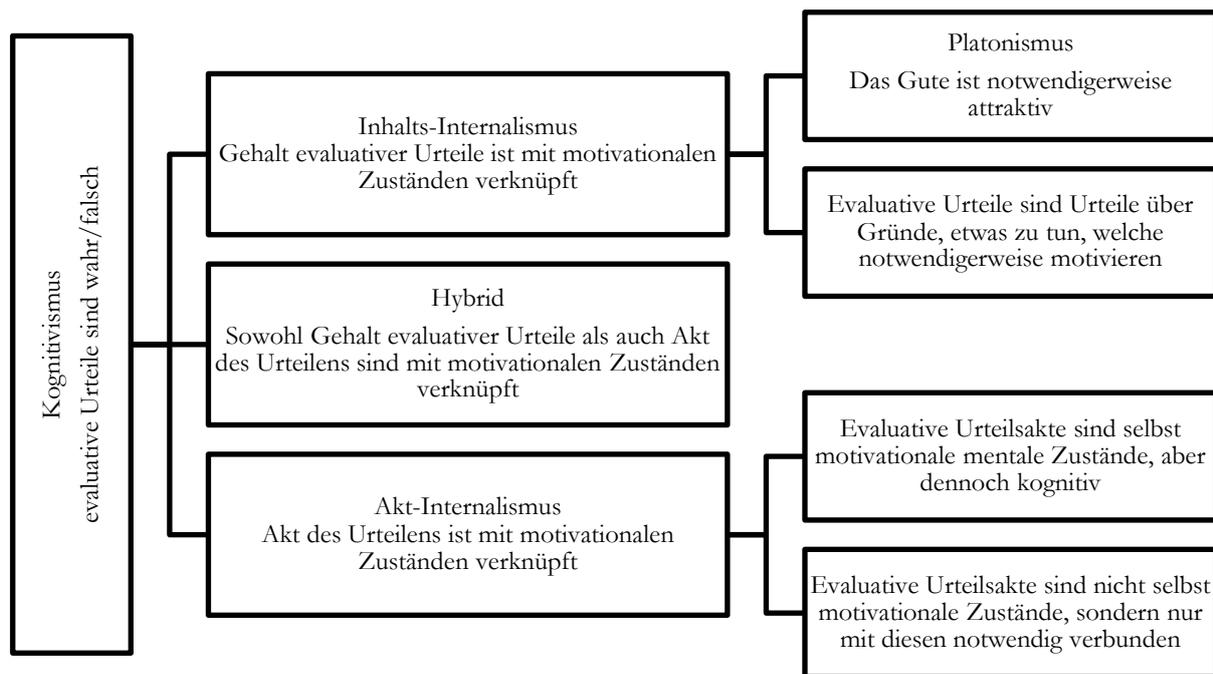
Kognitivisten können demnach entweder Prämisse 1 aufgeben und einen Urteils-Externalismus vertreten, oder Prämisse 2 ablehnen und damit eine alternative Theorie der Motivation entwickeln. In jedem Fall besteht die Aufgabe für Kognitivisten, den für Nonkognitivisten trivial bestehenden Zusammenhang zwischen Evaluation und Motivation zu erläutern.

2) Die Internalismus-These ist selbst in dem grundlegenden Schema bereits unklar, da sie keine Auskunft darüber gibt, welchen Aspekt der verschiedenen Elemente sie als notwendig miteinander verknüpft ansieht bzw. woran es liegt, dass diese beiden Elemente miteinander verknüpft sind: 1) den *Gehalt* von evaluativen (prudentiellen, moralischen etc.) Urteilen und der Motivation entsprechend zu handeln; 2) den *Urteilsakt* – das Fällen von, den Akt der Zustimmung zu, das Haben von solchen Urteilen oder äquivalente Akte – und der Motivation entsprechend zu handeln; 3) eine *Kombination* von beiden, also die Zustimmung zu einem spezifischen Gehalt und der Motivation, entsprechend zu handeln. In anderen Worten betrifft dies die Fragen: Liegt es an dem Akt des Urteilens bzw. der Zustimmung zu einem Urteil, dass Personen notwendig motiviert sind, entsprechend dieser Urteile zu handeln, an den Gehalten dieser Urteile – dass diese Urteile davon handeln, wie die Personen handeln sollen oder was das Beste zu tun ist etc. – oder an einer Kombination von beidem, also der Akt der Zustimmung zu einem Urteil über etwas Praktisches oder, spezifischer, etwas moralisch oder prudentiell Gebotenes.<sup>294</sup> Es lassen sich demnach folgende Positionen voneinander unterscheiden:

---

<sup>293</sup> „Necessarily, if one sincerely judges an action right, then one is motivated to some extent to act in accordance with that judgment. (*Motivational Judgment Internalism*) When taken by themselves, beliefs neither motivate nor generate any motivationally efficacious states. (*Motivational Humeanism*) Therefore, moral judgments are not beliefs. (*Moral Non-cognitivism*)” (Shafer-Landau 2003, 121). Vgl. auch Rosati 2008.

<sup>294</sup> Donald Davidson vertritt dieser Unterscheidung nach einen Internalismus, der den Akt der Urteilens betrifft, nicht den besonderen semantischen Gehalt des Geurteilten, während bei Richard Hare eine Theorie vorliegt, die den besonderen semantischen Gehalt der Urteile mit motivationalen Zuständen verknüpft. Aber: Vgl. Jordan Sobel 2001, der argumentiert, dass Hare kein ‚content internalist‘ sein kann, da die Urteile wesentlich keinen Inhalt (content) hätten.



David Brink unterscheidet in diesem Sinne zwischen ‚agent internalism‘, ‚appraiser internalism‘ und ‚hybrid internalism‘.<sup>295</sup>

„*Agent internalism* claims it is in virtue of the concept of morality that *moral obligations* motivate, or provide reason for, the agent to do the moral thing. Thus, it is a conceptual truth about morality, according to agent internalism, that agents have reason or motive to comply with their moral obligations. *Appraiser internalism* claims it is in virtue of the concept of morality that *moral belief* or *moral judgment* provides the appraiser with motivation or reason for action. Thus, it is a conceptual truth about morality, according to appraiser internalism, that someone who holds a moral belief or makes a moral judgment is motivated to, or has reason to, perform the action judged favorably. [...] *Hybrid internalism* claims it is a conceptual truth about morality that the *recognition of a moral obligation* motivates or provides the agent (the person who recognizes his obligation) with reason for action.“ (Brink 1989, 40f.)

Nach der erste Form liegt es demnach an der besonderen Natur der Gehalte, in diesem Fall moralischer Verpflichtungen, dass die Zustimmung von Personen zu diesen impliziert, dass sie motiviert sind, entsprechend zu handeln, während für die zweite Form es an der besonderen Natur des moralischen Urteilens liegt, dass diese mit Motivationen, entsprechend zu handeln, verbunden sind.<sup>296</sup>

Als Beispiele für Theorien, welche die Gehalte von evaluativen Urteilen mit der Motivation, entsprechend zu handeln, verknüpft sehen, könnte ein platonische Theorie dienen: das Gute ist so beschaffen, dass derjenige, der es kennt, es anstrebt, weil es gut ist. Alternativ dazu kann die Verbindung zwischen dem Gehalt evaluativer Urteile und der Motivation, entsprechend zu handeln, über den Begriff des praktischen Grundes expliziert werden. Ein Beispiel für eine

<sup>295</sup> David Brink behandelt hier Internalismus bezogen auf Urteile (hier ‚motives‘) und Internalismus bezogen auf Gründe.

<sup>296</sup> Vgl. auch Smith 1994, 61 mit der Unterscheidung zwischen ‚rationalism‘ und dem ‚practicality requirement on moral judgement‘. Vgl. auch Darwall 1992 und seine Unterscheidung zwischen ‚constitutive‘ und ‚non-constitutive internalism‘.

derartige Analyse findet sich bei Robert Dunn.<sup>297</sup> Basale evaluative Ausdrücke wie ‚ist gut‘, ‚ist richtig‘ und ‚ist obligatorisch‘ (‚sollen‘ bzw. ‚ought‘) lassen sich analysieren als ‚von großem Wert‘, ‚von hinreichendem Wert‘ und ‚von größtem Wert‘. Da die Zuschreibung eines Wertes zugleich die Zuschreibung von etwas ist, das *für* den jeweiligen Gegenstand spricht, können die vorigen Zuschreibungen auch über den Begriff des Grundes ausgedrückt werden: also ‚einigen Grund‘, ‚hinreichenden Grund‘ und ‚meisten Grund‘.

„It is the fact that *of value* is the semantic common denominator of such evaluative terms as characteristically occur in evaluatives apt for expressing all-out present-tense summary evaluative judgements about one’s own action which fundamentally accounts for why such judgements are analysable as judgements about present reason for oneself to act. For to say that *value* attaches to something *x* is effectively to say that there is *something to be said* (or *adduced*, or *cited*) *for* (or *on the side of*, or *in favour of*, or *in support of*) *x*; and where *x* is something upon which reasons coherently can be said to bear – for example, desiring, wanting, intending, believing, having certain feelings, and *acting* – this comes to saying that there is *some reason for x*.” (Dunn 1987, 97)

Wenn jemand demnach z.B. urteilt ‚Es wäre gut, wenn ich jetzt einkaufen gehe‘, so lässt sich dies analysieren als ‚Es wäre von großem Wert, wenn ich jetzt einkaufen gehe‘, was sich wiederum analysieren lässt als ‚Ich habe einigen Grund, jetzt einkaufen zu gehen‘. Die Verbindung zu Motiven, entsprechend des evaluativen Urteils zu handeln, ergibt sich über eine internalistische Deutung von praktischen Gründen: Eine Person, die rational ist, ist motiviert, entsprechend ihrer Gründe zu handeln. Die Analyse des Urteils-Internalismus greift an diesem Punkt demnach auf eine Analyse normativer Gründe zurück. Für Christine Korsgaard ist eine grundlegende Anforderung an praktische Gründe, dass sie rationale Personen, die sie anerkennen, auch zum Handeln motivieren.

„Practical-reason claims, if they are really to present us with reasons for action, must be capable of motivating rational persons. I will call this the *internalism requirement*.” (Korsgaard 1986, 11)

Diese Bedingung an praktische Gründe wird gemeinhin als analytisch aufgefasst: ein praktischer Grund ist *definiert* als etwas, das eine rationale Person zu entsprechenden Handlungen motivieren kann, wenn sie ihn als solchen (an)erkennt. Differenzen ergeben sich in weiterer Folge in der Analyse dessen, was genau ein solcher Grund ist, wie er rationale Personen motivieren kann und was mit ‚rational‘ genau gemeint ist.<sup>298</sup>

Um das Vorige kurz zusammenzufassen: eine Form von motivationalem Urteils-Internalismus kann die Gehalte von evaluativen Urteilen über das eigene Handeln als notwendig verknüpft sehen mit der Motivation, entsprechend zu handeln. Als verknüpfendes Element fungiert dafür der Begriff des praktischen Grundes. Aus der Definition eines praktischen Grundes ergibt sich,

---

<sup>297</sup> Vgl. Dunn 1987, ch. 4. Vgl. auch eine ähnliche Analyse bei Pendlebury 2002.

<sup>298</sup> Die philosophische Diskussion dreht sich auch vorwiegend um moralische Urteile und ob moralische Verpflichtungen Gründe für Handlungen sind oder nicht unabhängig von der motivationalen Verfasstheit, also z.B. bereits vorhandenen Wünschen und Strebungen der Person, auf die diese zutreffen.

dass eine Person, die glaubt, einen praktischen Grund zu haben, motiviert ist, entsprechend zu handeln, *solange sie rational ist*. Das Argument kann man folgendermaßen verstehen:<sup>299</sup>

- (P1) Wenn eine Person glaubt (oder urteilt), dass sie X tun sollte (oder ‚dass es gut wäre, X zu tun‘; ‚dass es richtig wäre, X zu tun‘), dann glaubt sie, dass sie einen Grund hat, X zu tun.
- (P2) Wenn eine Person glaubt (oder urteilt), dass sie einen Grund hat, X zu tun, dann ist sie motiviert, X zu tun, oder sie ist praktisch irrational.
- (K) Wenn eine Person glaubt, dass sie X tun sollte, dann ist sie motiviert, X zu tun, oder sie ist praktisch irrational

Dieses Argument behauptet nicht, dass eine Person deswegen bereits einen guten Grund hat, auf eine Weise zu handeln, nur weil sie glaubt, dass sie einen solchen habe. Das Argument sagt nur, dass eine rationale Person motiviert ist, entsprechend der Gründe zu handeln, die sie selbst für die Situation als solche anerkennt. Dagegen lässt sich sicherlich einwenden, dass eine Person, die fälschlich glaubt, dass sie einen Grund hat, nicht unbedingt irrational ist, wenn sie nicht motiviert ist, entsprechend ihrem Grund zu handeln, da sie schließlich *tatsächlich* keinen Grund hat, so zu handeln. Dieses Problem hat Ähnlichkeit mit dem Problem uneigentlicher Willensschwäche, also einem Handeln wider besseren aber *falschen* Glauben, da auch in diesen Fällen nicht klar ist, ob die Person nun tadelnswert handelt oder nicht. Die Beurteilung dieser Möglichkeit hängt daran, wie (praktische) Rationalität verstanden wird und welches Gewicht dem evaluativen Urteil über die eigenen Handlungsgründe zukommt.<sup>300</sup>

Urteils-Internalismen, welche nicht den Gehalt evaluativer Urteile über das eigene Handeln, sondern den Begriff des evaluativen Urteilens mit der Motivation, entsprechend zu handeln, verknüpfen, können dies auf zumindest zwei verschiedene Arten tun. Zum einen können sie meinen, dass evaluative Urteile über das eigene Handeln selbst bereits motivationale Zustände sind. Für Nonkognitivisten ist dies normalerweise der Fall, wobei argumentativ oftmals auf Basis eines Internalismus für die nonkognitivistische Position argumentiert wird und nicht umgekehrt für den Internalismus auf Basis eines Nonkognitivismus. Für Kognitivisten ist diese Position möglich, indem kognitive Zustände z.B. im Sinne einer ‚Besire-Theorie‘ als motivationale Zustände verstanden werden.<sup>301</sup> Bestimmte Glaubenszustände sind demnach sowohl Glaubens-

---

<sup>299</sup> „Belief Internalism: Beliefs about reasons necessarily involve motivation. We cannot believe that we have a reason to do something without being motivated to do this thing.” (Parfit 1997, 104) Bei Parfit fehlt in dieser Formulierung die Einschränkung auf rationale Akteure. Vgl. auch „If an agent judges that she has reason to A in circumstances C, then she is motivated to A in C.” (Garrard & McNaughton 1998, 48)

<sup>300</sup> Diese Frage diskutiere ich näher in Kap. VI.1.

<sup>301</sup> Das Wort ‚besire‘ setzt sich zusammen aus ‚belief‘ und ‚desire‘, was die Mischung der beiden mentalen Zustände in einem einzigen Zustand andeuten soll. Vgl. Zangwill 2008.

zustände als auch Wunschzustände oder allgemeiner Pro-Einstellungen.<sup>302</sup> Glaubenszustände können auch als motivationale Zustände auf der Basis einer dispositionalen Theorie von Überzeugungen aufgefasst werden. Nach dieser Theorie sind alle Glaubenszustände motivationale Zustände, da sie Dispositionen zu bestimmten Verhaltensweisen sind.<sup>303</sup>

Zum anderen können Theorien evaluative Urteilsakte als kognitiv verstehen und nicht selbst als motivationale Zustände für Handlungen, aber als mit motivationalen Zustände streng verbunden. Das Urteil motiviert demnach nicht direkt die Handlung, sondern nur indirekt über den Umweg, zuerst einen von ihr verschiedenen motivationalen Zustand, eine Pro-Einstellung, also z.B. einen Wunsch, zu motivieren, der bzw. die dann eine entsprechende Handlung motiviert. So argumentiert Thomas Nagel für eine Unterscheidung zwischen *motivierten* und *unmotivierten* Wünschen, wobei erstere auch hinreichend durch reine Glaubensakte motiviert sein könnten.<sup>304</sup> Die Zuschreibung eines motivierten Wunsches einer handelnden Person für ihre Handlung erfolge rein als Konsequenz über die Zuschreibung eines entsprechenden Zieles, das die Person mit ihrer Handlung verfolge. Dieser Wunsch sei aber nicht notwendigerweise selbst unmotiviert oder durch andere (unmotivierte) Wünsche motiviert, sondern könnte auch die Folge von reinen Überzeugungen, wie z.B. moralischen Erwägungen der Person, sein.

„The assumption that a motivating desire underlies every intentional act depends, I believe, on a confusion between two sorts of desires, motivated and unmotivated. [...] The claim that a desire underlies every act is true only if desires are taken to include motivated as well as unmotivated desires, and it is true only in the sense that *whatever* may be the motivation for someone's intentional pursuit of a goal, it becomes in virtue of his pursuit *ipso facto* appropriate to ascribe to him a desire for that goal. But if the desire is a motivated one, the explanation of it will be the same as the explanation of his pursuit, and it is my no means obvious that a desire must enter into this further explanation.” (Nagel 1970, 29)

3) Auch das Konsequenz des Konditionals, also die Seite der Motivation, wird unterschiedlich expliziert, je nachdem, was jeweils als motivationaler Zustand verstanden wird. ‚Humeaner‘ gehen davon aus, dass einzig in einem weiten Sinne Wünsche oder Pro-Einstellungen motivationale Zustände sein können und dabei streng von glaubensartigen Zuständen unterschieden seien.<sup>305</sup> Dies wird unter anderem mit dem Argument begründet, dass diese beiden mentalen Grundtypen verschiedene Passrichtungen (direction of fit) besäßen:<sup>306</sup> Während ein Glaube die Passrichtung

---

<sup>302</sup> Vgl. Smith 1994, ch. 4.7.

<sup>303</sup> Vgl. Bromwich 2010.

<sup>304</sup> Vgl. Nagel 1970, ch. V.

<sup>305</sup> In der Debatte werden oftmals humaneische und anti-humaneische Positionen voneinander unterschieden, meist ohne Anspruch auf historisch-sachliche Korrektheit sondern als vage Anknüpfung an die Philosophie Humes. Diese Form der Bezeichnung ist problematisch, auch wenn sachlich einiges dafür spricht, diese Positionen mit Hume in Verbindung zu bringen, wie z.B. in folgendem Zitat: „Since reason alone can never produce any action, or give rise to volition, I infer, that the same faculty is as incapable of preventing volition, or of disputing the preference with any passion or emotion.“ (Hume 1739, II.3.3 414f.)

<sup>306</sup> „The basic idea is that mental states can have one of two possible “directions of fit”. Beliefs are the exemplars of mental states with "mind-to-world" direction of fit since they aim at being true, i.e., "matching the world". If there is

Geist-zu-Welt besitze, habe ein Wunsch die Passrichtung Welt-zu-Geist. Dies heißt: Stimmt ein Glaube nicht mit der Welt überein, wird (rationalerweise) der Glaube aufgegeben. Stimmt ein Wunsch nicht mit der Welt überein, so wird (rationalerweise) der Wunsch nicht aufgegeben, sondern die Welt geändert. Umgekehrt wird ein Glaube nicht aufgegeben, wenn die Welt mit ihm übereinstimmt, während ein Wunsch aufgegeben wird, wenn die Welt so ist, wie sie gewünscht wird. Zusätzlich wird noch angenommen, dass es keine mentalen Zustände gebe, die beide Passrichtungen besitzen.<sup>307</sup> Weiters wird nun angenommen, dass mentale Zustände mit der Passrichtung Welt-zu-Geist, also in einem weiten Sinne Pro-Einstellungen, notwendig sind um Handlungen zu motivieren.<sup>308</sup> Für Handlungstheorien, die das Wunsch-Überzeugungs-Modell akzeptieren, ist dies selbstverständlich, es macht gewissermaßen den Kern einer solchen Theorie aus. Um die Notwendigkeit der Präsenz einer Pro-Einstellung nicht bloß zu stipulieren, sondern zu begründen, wird auf den teleologischen Charakter von Handlungserklärungen Bezug genommen, wie Michael Smith in seinem Argument:<sup>309</sup>

„(a) Having a motivating reason is, *inter alia*, having a goal (b) Having a goal is being in a state with which the world must fit and (c) Being in a state with which the world must fit is desiring.” (Smith 1994, 116)

Dieses Argument vermag jedoch noch nicht von sich aus eine Hume'sche Auffassung von motivierenden Gründen zu etablieren, da noch nicht klar ist, woher Pro-Einstellungen kommen und wodurch diese motiviert sind. So könnte Thomas Nagel, als Rationalist, dem Argument zustimmen, mit der Betonung, dass es eine Klasse von motivationalen Zuständen, nämlich motivierten Wünschen gebe, die selbst nicht wieder durch andere unmotivierte Wünsche motiviert sind. Es braucht demnach eine substantielle Auffassung darüber, was Wünsche genauer sind, die über die bloß formale Charakterisierung mittels der Passrichtung hinausgeht.<sup>310</sup> Oder das Argument wird durch ein weiteres formales Prinzip ergänzt wie z.B. das Prinzip des ‚desire-in-desire-out‘.<sup>311</sup> Wünsche oder allgemeiner Pro-Einstellungen sind nach diesem Prinzip nur dann Ergebnis von kognitiv-deliberativen Prozessen, wenn diese als *Input* zumindest einen Wunsch oder eine Pro-Einstellung haben, von denen sie abgeleitet werden können. Dies heißt, das

---

a mismatch between a belief and reality then it is the belief which has failed; it has failed to "fit" the world. Desires, on the other side, are exemplars of "world-to-mind" direction of fit. They aim at satisfaction or realization, not truth. If a desire fails to "fit" the world, that is not any defect in the desire. It merely shows that the world needs to be changed if the desire is to be realized.” (Schueler 1991, 277) Vgl. Anscombe 1963; Smith 1994; Searle 1983; Humberstone 1992. Für Kritik vgl. Sobel & Copp 2001; Little 1997.

<sup>307</sup> Vgl. Smith 1994, 111ff.

<sup>308</sup> Die internalistische These lautet dann genauer: Wenn [evaluatives Urteil ( $\varphi$ )], dann [Pro-Einstellung ( $\varphi$ )]

<sup>309</sup> Auf ähnliche Weise hat bereits Bernard Williams gegen die Existenz externer Gründe argumentiert. Vgl. Williams 1979; Williams 1989.

<sup>310</sup> Vgl. dazu Nagel 1970; Wallace 1990; Korsgaard 1986. Nach Heath 1997 liegt der Diskussion zwischen Rationalisten und Humeanern bezüglich praktischer Vernunft eine falsche Auffassung von epistemischer Rechtfertigung, nämlich eine fundamentalistische, zugrunde.

<sup>311</sup> So genannt von Jay Wallace in Wallace 1990.

Argument der Passrichtung wird nochmals angewendet, aber nicht auf motivierende Gründe für Handlungen, sondern auf motivierende Gründe für Pro-Einstellungen bzw. Wünsche.<sup>312</sup> Außerdem ist unklar, weshalb die mentalen Zustände, die inter alia angenommen werden, nicht mit gleichem Recht motivationale Zustände genannt werden sollten. Dies betrifft mithin die Frage, weshalb einzig die Pro-Einstellung bzw. der Wunsch die ‚motivationale Arbeit‘ verrichtet und nicht ebenso die anderen mentalen Zustände, die offensichtlich nötig sind, damit eine Person zum Handeln motiviert ist.<sup>313</sup>

Die Signifikanz des Passrichtungs-Arguments für die motivationale Relevanz von evaluativen Urteilen ist strittig. Für Nonkognitivisten ist es ein weiteres Argument dafür, dass evaluative Urteile nicht wahrheitswertfähig sind, da sie *konative* Zustände statt *kognitiver* Zustände ausdrücken. Für Kognitivisten sind evaluative Urteile wahrheitswertfähig, d.h. sie drücken Glaubenszustände mit propositionalem Gehalt aus, die wahr oder falsch sein können. Diese werden gemäß der Passrichtung Geist-zu-Welt aufgegeben, wenn sie falsch sind, also gemäß einer korrespondenztheoretischen Auffassung von Wahrheit nicht mit der Welt übereinstimmen. Wenn ich glaube, dass P, und feststelle, dass non-P, dann höre ich auf zu glauben, dass P, wenn ich rational bin. Wenn es falsch ist, dass man keine falschen Versprechen abgeben soll, dann höre ich auf zu glauben, dass man keine falschen Versprechen abgeben soll. Andererseits ist es kein guter Grund, ein evaluatives Urteil aufzugeben, nur weil die Welt sich nicht so verhält, wie es das Urteil aussagt. Wenn einige Personen falsche Versprechen abgeben heißt dies nicht, dass es erlaubt ist, falsche Versprechen abzugeben oder gar dass es geboten wäre, falsche Versprechen abzugeben. In diesem Sinne ändert sich mein Urteil nicht aufgrund einer Nicht-Übereinstimmung mit der Welt.

Folgende Reaktionen auf das Passrichtungs-Argument sind unter anderem möglich: 1) Die Sinnhaftigkeit und Konsistenz der Unterteilung mentaler Zustände anhand der Passrichtung wird allgemein bestritten; 2) Man nimmt mentale Zustände an, die beide Passrichtungen besitzen (besires); 3) Das Argument wird anerkannt, aber man betont, dass reine Glaubenszustände Pro-Einstellungen motivieren können – der Ursprung der Motivation sind mithin die Glaubenszustände und nicht die Pro-Einstellungen; 4) Man versteht Glaubenszustände als motivationale Zustände, die Handlungen alleine erklären können. Klar ist, dass die Frage, was genau ein motivationaler Zustand eigentlich ist, alles andere als einheitlich beantwortet wird. Die Debatte dreht sich mithin nicht nur um die Frage, wie genau das Verhältnis zwischen motivierenden

---

<sup>312</sup> Vgl. „For if a desire is motivated then it is susceptible to teleological explanation. The motive that motivates the desire must therefore embody some goal of the agent. But no belief could do that, for beliefs are not the right kind of state to embody the having of goals.” (Smith 1987–88, 251f.) Smith wiederholt dieses Argument *nicht* in Smith 1994. Weshalb die Erklärung für die Bildung eines Wunsches teleologisch sein muss, wird nicht weiter erläutert.

<sup>313</sup> Vgl. auch die Kritik von Rüdiger Bittner in Bittner 2011, Kap. 1–2; Kap. 9.

Gründen (Motiven) und normativen Gründen (guten Gründen) genau beschaffen ist, sondern auch um jene, wie sich Motive und ‚motiviert sein, etwas zu tun‘ zueinander verhalten. Heißt ein Motiv haben dasselbe wie ‚motiviert sein, etwas zu tun‘? Und welche mentalen Zustände können die funktionale Rolle erfüllen, Handlungen zu motivieren?

In der Diskussion spielen neben unqualifizierten bzw. nicht-konditionalen Formen des Urteils-Internalismus vor allem konditionale bzw. qualifizierte Formen eine stärkere Rolle. Die notwendige Verbindung zwischen Urteilen und der Handlungsmotivation gilt demnach nur, wenn eine bestimmte Bedingung erfüllt ist.<sup>314</sup>

Notwendig: Wenn eine Person aufrichtig urteilt, dass sie F tun sollte, dann ist sie auch motiviert dazu, F zu tun, oder die Person ist X.

Für X werden bestimmte Vorschläge gemacht wie: ‚ist psychisch nicht normal‘, ‚ist praktisch irrational‘ oder ‚ist nicht tugendhaft‘.<sup>315</sup> Diese Vorschläge schwächen die internalistische These in je charakteristischer Form ab und ändern deren argumentative Stellung in der Debatte. Das Problem an diesen Abschwächungen ist generell, dass sich der vermutete Zusammenhang zwischen einem evaluativem Urteil und Handlungsmotivation nicht mehr nur dem evaluativen Urteil, sondern auch dem zusätzlichen Faktor zurechnen lässt. Der zusätzliche Faktor muss deshalb in einleuchtender Weise mit dem evaluativen Urteil in Beziehung stehen, da ansonsten die Motivation nicht mehr dem Urteil ‚intern‘ zugerechnet werden kann, sondern ‚extern‘ der Verknüpfung des Urteils mit einem ihr äußerlichen Umstand oder Faktor. Wenn diese *ceteris paribus* Bedingung<sup>316</sup> außerdem nicht näher spezifiziert wird, droht die Trivialisierung der internalistischen These: ‚Wenn (Urteil) dann (Motivation oder keine Motivation)‘ ist schließlich aus rein logischen Gründen wahr.

### III.5.2 Urteils-Internalismus, Proportionalismus und Willensschwäche

Die Frage ist nun, in welchem Verhältnis das Phänomen der Willensschwäche zum Urteils-Internalismus verschiedener Prägung steht. Willensschwäche wird gemeinhin als Gegenbeispiel für starke, nicht-konditionale Formen von Urteils-Internalismen angesehen. Dies deshalb, da bei Fällen von Willensschwäche ein evaluatives Urteil vorliegt, aber die Person nicht entsprechend motiviert ist, dem Urteil gemäß zu handeln. Michael Smith fügt daher in seine Formulierung des internalistischen Zusammenhanges Willensschwäche als Ausnahme hinzu:

---

<sup>314</sup> „If an agent judges that it is right for her to  $\phi$  in circumstances C, then either she is motivated to  $\phi$  in C or she is practically irrational.” (Smith 1994, 61) „*Conditional internalism*: Necessarily, if a person judges that she morally ought to  $\phi$ , then she is (at least somewhat) motivated to  $\phi$  if she is C.” (Björkland & al. 2012, 126) Schematisch lässt sich das wie folgt darstellen: Notwendig: Wenn [evaluatives/normatives Urteil ( $\varphi$ )], dann [entweder (Motivation ( $\varphi$ )) oder (die Person ist X)].

<sup>315</sup> Vgl. Björkland & al. 2012, 127f.

<sup>316</sup> Vgl. die Argumentation für die Einfügung einer *ceteris paribus* Bedingung bei Kennett & Smith 1994.

„Weak Internalism: If an agent judges it right to  $\phi$  in certain circumstances C, then she is motivated to  $\phi$  in C, at least absent weakness of will and the like.” (Smith 1996, 175)

Doch weshalb sollte Willensschwäche überhaupt ein Gegenbeispiel für starke, nicht-konditionale Formen des Internalismus sein? Ein Gegenbeispiel ist ein Fall, in dem die Bedingungen des Antezedens erfüllt sind, doch das Konsequenz ausbleibt. Dies wäre dann der Fall, wenn eine Person ein evaluatives Urteil fällt, aber dennoch in *keiner* Weise motiviert ist, entsprechend zu handeln. In der philosophischen Diskussion sind dies Fälle von Amoralisten und depressiven Personen.<sup>317</sup> Im Unterschied dazu wird normalerweise davon ausgegangen, dass willensschwache Personen motiviert sind, entsprechend ihrem evaluativen Urteil zu handeln, nur dass diese Motivation schwächer ist als die Motivation für die alternative, willensschwache Handlung. Die willensschwache Person befindet sich in einem charakteristischen Konflikt, der sich durch zwei widerstrebende Motive auszeichnet, nicht durch das vollständige Ausbleiben der Motivation.<sup>318</sup> Sowohl starke als auch schwache Formen des Urteils-Internalismus machen keine Aussage über die Stärke der Motivation, die mit dem evaluativen Urteil verbunden ist, und sei sie noch so schwach. Diese Thesen sind in ihrer Grundformulierung daher nicht-proportionalistisch, d.h. sie formulieren einen Zusammenhang zwischen Evaluation und Motivation, ohne dabei die Güte der Evaluation mit der Stärke an Motivation zu korrelieren.<sup>319</sup> Es sollten daher proportionalistische von nicht-proportionalistischen Versionen von Internalismen unterschieden werden:

*Nicht-proportionaler Urteils-Internalismus:* Notwendig:<sup>320</sup> Wenn eine Person aufrichtig urteilt, dass sie F tun sollte, dann ist sie auch motiviert dazu, F zu tun, aber der Grad der Motivation ist nicht notwendigerweise proportional zur Güte der in dem Urteil anerkannten Gründe zu handeln oder der Glaubensstärke in das Urteil.

*Proportionaler Urteils-Internalismus:*<sup>321</sup> Notwendig: Wenn eine Person aufrichtig urteilt, dass sie F tun sollte, dann ist sie auch motiviert dazu, F zu tun, und der Grad der Motivation ist proportional zur Güte der in dem Urteil anerkannten Gründe zu handeln oder der Glaubensstärke in das Urteil.

---

<sup>317</sup> Zum Amoralisten vgl. Brink 1989, ch. 3.3. Zu Depression und ‚listlessness‘ vgl. Mele 1996; Tenenbaum 2003.

<sup>318</sup> „To give but one example, Smith tries to refute the anti-Humean who espouses internalism by pointing out that "practical irrationalities of various kinds – weakness of will and the like – can break the connection between moral judgment and motivation" ([Smith] 1994, p. 120). But of course such practical irrationalities only sever the connection between judgment and action. They do nothing to support the highly implausible claim that the weak-willed person is not at all motivated to act in accordance with her best judgment.” (Garrard & McNaughton 1998, 49)

<sup>319</sup> Zum Proportionalismus vgl. auch Mark Schroeder 2007, ch. 6; Zangwill 2008.

<sup>320</sup> „VB It is essential to our self-directed moral beliefs that we ought to do something that we have a desire to act in accordance with those beliefs of some non-zero strength, but what non-zero strength is not essential.” (Zangwill 2008, 53)

<sup>321</sup> Schematisch: Notwendig: Wenn [evaluatives/normatives Urteil ( $\varphi$ )], dann [proportionale Motivation ( $\varphi$ )] „PB It is essential to our self-directed moral beliefs that we ought to do something that we have a desire to act in accordance with those beliefs of a certain strength; in particular, the strength of that desire is proportional to the degree of the moral belief.” (Zangwill 2008, 53) Der Proportionalismus findet sich bereits bei Davidson, formuliert in seinem Prinzip 2, vgl. Kap. II.2.2.

Die proportionale Version des Urteils-Internalismus gerät nun in Konflikt mit dem Phänomen der Willensschwäche, wie sich anhand folgenden Arguments zeigen lässt:

- (P1) Wenn eine Person aufrichtig urteilt, dass sie besser F-en sollte als zu G-en, dann urteilt sie, dass sie mehr Gründe hat zu F-en als zu G-en (Analyse evaluativer Urteile in Gründe-Urteile)
- (P2) Wenn eine Person urteilt, dass sie mehr Gründe hat zu F-en als zu G-en, dann ist sie auch stärker motiviert zu F-en als zu G-en. (Internalismus und Proportionalismus)
- (P3) Wenn eine Person stärker motiviert ist zu F-en als zu G-en, dann F-t sie, wenn sie etwas intentional tut<sup>322</sup>
- (K1) Wenn eine Person aufrichtig urteilt, dass sie besser F-en sollte als zu G-en, dann F-t sie, wenn sie etwas intentional tut.
- (K2) Aber: Eine willensschwache Person urteilt aufrichtig, dass sie besser F-en sollte als G-en, und G-t dennoch intentional

Willensschwäche ist also in seinem normalen Verständnis nur dann ein potentiell Gegenbeispiel für starke Formen des Urteils-Internalismus, wenn dieser durch eine Proportionalismus-These ergänzt wird, d.h. die Idee einer Entsprechung zwischen der in dem evaluativen Urteil ausgedrückten Güte der Gründe und der motivationalen Stärke, entsprechend zu handeln.<sup>323</sup> Diese Ergänzung ist nötig, da die willensschwache Person zumindest in irgendeinem positiven Grad motiviert ist, entsprechend ihrem besseren Urteil zu handeln, und damit der nicht-proportionalistischen Version des Urteils-Internalismus genügt. Dies führt zu der Frage, ob auch etwas gegen nicht-proportionale Versionen des Internalismus spricht. Internalistische Theorien erläutern den handlungsleitenden Charakter von evaluativen Urteilen als *interne* Eigenschaft solcher Urteile. Internalistische Theorien sollen daher verständlich machen, weshalb Personen normalerweise motiviert sind, ihre evaluativen Urteile in Handlungen umzusetzen, und dies anhand von Eigenschaften der evaluativen Urteile selbst erläutern. Da nicht-proportionale Versionen jedoch keine Auskunft über die zu erwartete Stärke der Motivation geben, können sie auch nichts über die zu erwartende Handlung von Personen aussagen, die zu mehr als nur einer Handlung motiviert sind. In präziseren Worten: Sei J die Menge an normativen Urteilen einer Person über alternative Handlungen und M die Menge an motivationalen Zuständen derselben Person. Der Proportionalismus nimmt an, dass beide Mengen geordnet sind und dass alle

---

<sup>322</sup> Prämisse 3 dieses Argumentes ist angelehnt an das erste Prinzip von Davidson. Vgl. Kap. II.2.2.

<sup>323</sup> Dasselbe Argument lässt sich auch für einen Proportionalismus bezüglich der Glaubensstärke formulieren: (P1) Wenn eine Person aufrichtig urteilt, dass sie besser F tun sollte als G, dann ist sie stärker davon überzeugt, F zu tun als G zu tun (P2) Wenn eine Person stärker davon überzeugt ist, F zu tun als G zu tun, dann ist sie auch stärker motiviert, F zu tun als G zu tun (Internalismus & Proportionalismus) (P3) Wenn eine Person stärker motiviert ist, F zu tun als G zu tun, dann tut sie F, wenn sie etwas intentional tut (K) Aber: Eine willensschwache Person urteilt aufrichtig, dass sie besser F tun sollte als G, und tut dennoch G intentional.

Elemente in J ohne Informationsverlust auf M abgebildet werden können, d.h. ohne Verlust der Ordnungsrelationen in J. Sei  $f(x)$  eine solche Abbildung von J auf M (also  $f: J \rightarrow M$ ). Wenn nun der Grad der Überzeugung in ein normatives Urteil  $j_1$  größer ist als in ein davon verschiedenes normatives Urteil  $j_2$  (also  $j_1 > j_2$ ), dann ist  $f(j_1)$  auch größer als  $f(j_2)$ . Dagegen nehmen nicht-proportionalistische Internalismen nur an, dass alle Elemente von J auf M abgebildet werden können, ohne die Struktur dieser Abbildung näher zu bestimmen. So könnten im Extremfall alle Elemente von J auf dasselbe Element M abgebildet werden, was bedeuten würde, dass jedem normativen Urteil genau ein motivationaler Zustand entspricht. Einige einfache abstrakte Beispiele sollen die Probleme mit nicht-proportionalen Versionen von Internalismen aufzeigen.<sup>324</sup>

*Situation 1:* Zu Zeitpunkt  $t_1$  glaubt eine Person X sehr stark, dass sie F tun sollte. Zu einem späteren Zeitpunkt  $t_2$  glaubt X nur sehr schwach, dass sie F tun sollte.

Gemäß proportionalistischen Urteils-Internalismen ist die Motivation von X, F zu tun, zu  $t_2$  *geringer* als zu  $t_1$ . Diese Erwartung ist nicht gerechtfertigt durch nicht-proportionalistische Internalismen: X könnte zu  $t_2$  weniger, mehr oder gleich motiviert sein, F zu tun als zu  $t_1$ .

*Situation 2:* Zu Zeitpunkt  $t_1$  urteilt Person X, dass sie mehr Gründe hat F zu tun. Zu einem späteren Zeitpunkt  $t_2$  urteilt X, dass sie mehr Gründe hat F zu tun als zu  $t_1$ .

Normalerweise wird erwartet, dass X zu  $t_2$  mehr motiviert ist, F zu tun als zu  $t_1$ . Aber auch diese Erwartung ergibt sich nicht aus dem nicht-proportionalen Internalismus, da dieser nur aussagt, dass X zu beiden Zeitpunkten motiviert ist, F zu tun, nicht aber, in welchem Maße.

*Situation 3:* Zu Zeitpunkt  $t_1$  urteilt Person X, dass sie mehr Gründe hat F zu tun als G zu tun. Zu einem späteren Zeitpunkt  $t_2$  urteilt X, dass sie weniger Gründe hat F zu tun als G zu tun.

Normalerweise wird erwartet, dass sich die Motivation der Person in dieser Situation umkehrt. Diese Erwartung folgt jedoch nur aus proportionalistischen Versionen von Internalismen.

Situation 1 zeigt, dass auf Basis eines nicht-proportionalistischen Internalismus normative Urteile nicht handlungsleitend sind in dem Sinne, dass die Motivation für eine Handlung sensitiv ist gegenüber Änderungen der Glaubensstärke von normativen Urteilen über eine einzelne Handlung. Situation 2 zeigt dasselbe anhand der Stärke der Gründe für eine einzelne Handlung. Situation 3 schließlich zeigt dies anhand der Änderung von komparativen normativen Urteilen gegenüber zweier alternativer Handlungen. Nicht-proportionalistische Urteils-Internalismen sind demnach nicht sensitiv gegenüber Änderungen der normativen Bewertung einer Person über die Zeit hinweg. Daher sind diese Varianten nicht geeignet, um die handlungsleitende Funktion normativer und evaluativer Urteile adäquat auszudrücken. Für wahre Internalisten können nur Eigenschaften der normativen Urteile selbst deren handlungsleitende Kraft erläutern, da

---

<sup>324</sup> X steht dabei für Personen, t für Zeitpunkte, F und G für Typen von Handlungen.

ansonsten die motivationale Kraft einem dem normativen Urteil externen Faktor zuzuschreiben ist. Diese Überlegungen zeigen ein Dilemma für internalistische Theorien auf: nicht-proportionale Versionen scheitern an ihrem eigenen Anspruch, den handlungsleitenden Charakter von evaluativen Urteilen mittels der internalistischen These zu erläutern, proportionale Versionen haben in ihrer starken Form entweder die Schwierigkeit, Willensschwäche zu integrieren, oder in ihrer schwachen Form das Problem, die motivationale Kraft von Urteilen nicht alleine an internen Eigenschaften des Urteils selbst festzumachen.

### III.5.3 Sergio Tenenbaum: Appearances of the good

Sergio Tenenbaum verteidigt in seinem Buch *Appearances of the good* (2007) eine scholastische Sicht des Verhältnisses von Evaluation und Motivation, d.h. eine Form von proportionalistischen Urteils-Internalismus.<sup>325</sup>

„The version of the scholastic view I defend in this book is a particularly strong one. Desiring is, in my view, simply *identified* with conceiving something to be good from a certain perspective. [...] A consequence of this version of the scholastic view is that there can be no gap between evaluation and motivation; we cannot strongly desire what we conceive as only a little bit good or be only weakly motivated to pursue what we judge to be to a large extent good.” (Tenenbaum 2007, 14f.)

Genauer unterscheidet Tenenbaum in seinem Artikel *Accidie, evaluation, and motivation* (2007b) zwischen vier verschiedenen Aussagen über das Verhältnis von Evaluation und Motivation:

- „(S1) If  $\alpha$  desires X, then  $\alpha$  conceives X to be good. [...]
- (S2) If  $\alpha$  conceives X to be good, then  $\alpha$  desires X. [...]
- (S3)  $\alpha$  desires X if and only if  $\alpha$  conceives X to be good. [...]
- (S4) To desire X is to conceive X to be good.” (Tenenbaum 2003, 148)

S2 ist eine Variante eines Urteils-Internalismus, S1 formuliert ein Verständnis von ‚wünschen‘ als evaluative Einstellung. S3 ist die Konjunktion von S1 und S2, und S4 formuliert eine Identität der mentalen Einstellungen des Wünschens und des Als-Gut-Auffassens, die auch Tenenbaum selbst vertritt. Der Urteils-Internalismus wird dabei qualifiziert durch zwei Abschwächungen: *erstens* wird ‚etwas wünschen‘ identifiziert damit, dass etwas einer Person als gut erscheint, nicht, dass es tatsächlich gut ist; *zweitens* erscheinen Sachverhalte als gut innerhalb einer *Perspektive* auf den Sachverhalt, die nicht unbedingt eine reflektierte und endgültige Sicht der Person auf den Sachverhalt widerspiegeln muss. Die erste Abschwächung spricht gegen die Bezeichnung dieser Position als *Urteils-Internalismus*, schließlich besteht in dieser Variante die notwendige Verbindung nicht zwischen evaluativen Urteilen und der Motivation, entsprechend zu handeln, sondern zwischen Erscheinungen von etwas als gut und der Motivation, es anzustreben.

---

<sup>325</sup> Als ‚scholastisch‘ bezeichnet Tenenbaum alle Positionen, welche eine starke bzw. enge Verbindung von Motivation und Evaluation vertreten.

Insbesondere soll dies ermöglichen, dass eine Person motiviert sein kann, etwas zu tun, von dem sie weiß, dass es nicht gut ist. Die Bezeichnung als Urteils-Internalismus ist dennoch gerechtfertigt, da Tenenbaum unbedingte evaluative Urteile mit handlungsleitenden Intentionen, etwas zu tun, identifiziert.<sup>326</sup>

„Intentions are most naturally seen in the scholastic view as unconditional evaluative judgments that either are embodied in or precede action. Just as in the case of desire, we should not think of the judgment as something other than the intention. And here, too, we identify evaluation and motivation by saying that the judgment is the intention.“ (Tenenbaum 2007, 66)

Proportionalistisch ist dieser Internalismus, da er die Stärke der Motivation mit der Güte des Gehalts der evaluativen Urteile (der Stärke der Gründe oder der Stärke des Urteilsaktes) verbindet. Leider schwankt Tenenbaum gelegentlich zwischen einer normativen und einer faktischen Interpretation dieses Zusammenhangs, was die Interpretation erheblich erschwert. Einschränkungen der Gültigkeit oder Anwendbarkeit entsprechender Aussagen auf rationale Akteure sprechen für eine normative Interpretation, während die Identifikation von ‚wünschen‘ (desiring) mit ‚etwas als gut ansehen‘ (conceiving to be good) jedoch grundsätzlich für eine faktische Interpretation spricht. Die zweite Abschwächung soll die Existenz motivationaler bzw. evaluativer Konflikte innerhalb einer Person ermöglichen: was aus einer Perspektive als gut erscheint, kann aus anderen Perspektiven als schlecht bzw. weniger gut erscheinen. Perspektiven sind relativ gegenüber einer ihnen übergeordneten Perspektive, einer Konzeption des Guten der jeweiligen Person. So werden Erscheinungen des Guten, also Wünsche, anhand dieser Konzeption eines objektiv Guten aus Sicht der Person verglichen und gegebenenfalls korrigiert oder als Täuschung entlarvt.

Akrasia bzw. Willensschwäche stellt diese Konzeption vor das bereits bekannte Problem, dass bei Willensschwäche auf den ersten Blick Evaluation und Motivation auseinandergehen.<sup>327</sup> Aufgrund der Identifikation von Wünschen mit evaluativen Einstellungen deutet Tenenbaum Willensschwäche letztlich als kognitives Defizit.<sup>328</sup> Dafür überträgt er die Unterscheidung von Descartes zwischen direkten und mittelbaren Erkenntnissen (direct and oblique cognition) auf den Bereich praktischer Rationalität.<sup>329</sup> Wünsche bzw. Erscheinungen des Guten präsentieren

---

<sup>326</sup> Tenenbaum verweist hier auf Davidson mit seiner Unterscheidung zwischen bedingten und unbedingten evaluativen Urteilen und der Gleichsetzung letzterer mit Intentionen. Vgl. Kap. II.2; Davidson 1969a.

<sup>327</sup> „The scholastic identification of desiring and conceiving to be good, of motivational and evaluative attitudes, faces a challenge from cases of akrasia. In cases of akrasia, agents are not motivated in accordance with comparative evaluations. An akratic agent will think that A is better than B yet pursue B. In these cases, the agent’s motivational states seem not to fall in line with his evaluative states.“ (Tenenbaum 2007, 257)

<sup>328</sup> Vgl. auch Tenenbaum 1999.

<sup>329</sup> „We can say, roughly, a direct cognition is a representation of an object or claim through which one clearly understands (or seems to understand) *why* the object is as one represents it, or *how* it is that this claim is true. An oblique cognition is a representation of a claim or an object which is not a direct cognition, but one through which one understands (or seems to understand) *that* there are reasons to accept that the object is as one represents it, or

ihre jeweiligen Objekte direkt als gut unter der jeweiligen Perspektive, während dieselben Objekte in Alles-in-allem-Urteilen nur mehr indirekt repräsentiert werden: Die Person wisse zwar, *dass* etwas für die eine Handlung und gegen die Ausführung der anderen Handlung spreche, aber sie ist nicht mehr direkt in Kontakt mit den Gründen, welche für dieses Urteil sprechen. Im Übergang von bedingten evaluativen Urteilen zu Alles-in-allem-Urteilen verliert die Person die Gründe, die für die einzelnen Urteile sprechen, aus den Augen. Aus dieser Differenz von unmittelbar einleuchtenden evaluativen Auffassungen und abstrakten reflexiven Urteilen entspringt die Möglichkeit der Willensschwäche, die analog zu Davidson als Bildung eines unbedingten Urteils wider das als besser erachtete Alles-in-allem-Urteil verstanden wird.

„Turning back to the akratic person, we may say similarly that the akratic agent’s understanding that *A* is better than *B* is a reflective appearance of which she might have only a vague or oblique understanding. And in the same way that someone who has read the First Meditation might believe that he ought to be persuaded by the argument while not actually being persuaded by it, the akratic agent believes that he ought to be persuaded by his reflective understanding but is not persuaded by it. And this amounts to saying that the akratic agent acts against his own recognition that all things considered he should act otherwise.” (Tenenbaum 2007, 277)

Problematisch an diesem Konzept ist *erstens* die Identifikation des mentalen Zustandes des Wunsches mit dem Haben einer evaluativen Einstellung.<sup>330</sup> Dies schränkt das Spektrum von Wünschen auf rationale oder auf das Gute gerichtete Wünsche ein und ist daher nicht anwendbar auf andere Lebewesen und kleine Kinder, da diese nicht über die dafür notwendigen kognitiven und begrifflichen Fähigkeiten verfügen.<sup>331</sup> *Zweitens* ist das Verständnis des Alles-in-allem-Urteils problematisch. Nach Tenenbaum sind der Person, die ein solches Urteil fällt, die einzelnen Gründe, die für und wider die verschiedenen Handlungsoptionen sprechen, nicht mehr direkt präsent, d.h. nur mittelbar (oblique) gegeben. Das Alles-in-allem-Urteil ist aber gerade eine Abwägung aller Gründe, die für und wider eine Handlungsoption sprechen. Dies ist nicht möglich, wenn im Übergang von den bedingten Einzelurteilen zum abstrakten Alles-in-allem-Urteil genau diese Gründe nicht mehr präsent sind, die Person also nur mehr weiß, *dass* etwas für oder gegen eine Option spricht, aber nicht mehr, *was*. *Drittens* setzt das Konzept auf eine künstliche Unterscheidung zwischen evaluativen Auffassungen (conceivings) und evaluativen Urteilen. Während evaluative Auffassungen identisch sind mit Wünschen, bleibt der Status von evaluativen Urteilen unklar, außer, es handelt sich um unbedingte Urteile, welche mit Intentionen gleichgesetzt werden. Mithin bleiben die Bedenken auch gegenüber dieser Version des deskriptiven Urteils-Internalismus bestehen.

---

that the claim is true. We can think of an oblique cognition as standing proxy to one or more direct cognitions. The direct cognitions carries [sic] with it the explicit justification of that which the cognition represents as true.” (Tenenbaum 1999, 894)

<sup>330</sup> Zu diesem Einwand vgl. Kap. III.5.4.

<sup>331</sup> Vgl. Tenenbaum 2007, ch. 6.

### III.5.4 Urteils-Externalismus und Willensschwäche

Deskriptive Urteils-Externalismen verneinen die Existenz einer nicht-trivialen notwendigen Verbindung zwischen evaluativen oder normativen Urteilen und motivationalen Zuständen einer Person.<sup>332</sup> Dies ist nicht zu verwechseln mit der Behauptung des konträren Gegenteils, also dass Personen *niemals* motiviert wären, entsprechend ihrer normativen Urteile zu handeln. Es besagt lediglich, dass der Grund für die Entsprechung zwischen Evaluation und Motivation keine Eigenschaft des normativen Urteils alleine (intern), sondern durch einen weiteren, den Urteilen externen Faktor bedingt ist. Dieser Faktor ist die Pro-Einstellung der Person gegenüber dem Sachverhalt des evaluativen Urteils. Nach dieser Sicht reicht es demnach nicht aus, zu glauben, dass eine bestimmte Handlung besser zu tun wäre als eine andere, um im entsprechenden Maß motiviert zu sein, sie auch tatsächlich zu tun. Die Person muss außerdem eine Pro-Einstellung gegenüber der Handlung einnehmen und letztlich eine Intention bilden, die Handlung auch tatsächlich auszuführen. Eine einfache Formulierung eines motivationalen Urteils-Externalismus lautet:

*Motivationaler Urteils-Externalismus:* Wenn eine Person aufrichtig urteilt, dass sie F-en sollte, und eine Pro-Einstellung gegenüber F-en besitzt, dann ist die Person motiviert zu F-en.

Die Formulierung drückt eine relative Unabhängigkeit des motivationalen Systems von Personen gegenüber ihrem evaluativen Überzeugungssystem aus. Zur Handlungsmotivation sind nach dieser Auffassung zumindest zwei Komponenten notwendig, eine kognitiv-evaluative und eine konative Komponente. Ohne ein Verständnis der konativen, strebenden Komponente und allgemein des motivationalen Systems bleibt dieser Zusammenhang noch recht abstrakt und grob. In einem sehr weiten Verständnis von Motivation kann jeder mentale Akt andere Akte motivieren.

„**Motivation** ist der allgemeine Begriff für alle Prozesse, die der Initiierung, der Richtungsgebung und der Aufrechterhaltung physischer und psychischer Aktivitäten dienen.“ (Gerrig & Zimbardo 2008, 414)

Die Wahrnehmung eines Fotos mag die Erinnerung an einen Urlaub motivieren, der Gedanke an die morgige Prüfung mag die Sorge motivieren, diese nicht zu bestehen, die Vorstellung eines Apfels mag den Wunsch motivieren, etwas zu essen. Eine der grundlegendsten Mechanismen des Geistes ist dabei sicherlich die Assoziation.<sup>333</sup> In einem engeren Sinn umfasst der Begriff der Motivation diejenigen Zustände, welche *Handlungen* motivieren. Philosophische Theorien zur Willensschwäche unterscheiden sich grundlegend darin, ob sie als Quellen der Handlungsmotivation konative und affektive Zustände wie Emotionen, Wünsche und Affekte annehmen,

---

<sup>332</sup> Als trivial verstehe ich hier Formulierungen von Internalismen mit *ceteris paribus* Klauseln, die nicht weiter spezifiziert werden. Diese sind aus rein logischen Gründen wahr, aber daher auch inhaltlich leer.

<sup>333</sup> Vgl. Lear 2009.

oder darüber hinaus auch rein kognitive Zustände wie Überzeugungen, Vorstellungen, Wahrnehmungen und Phantasien. Die humane Theorie der Motivation nimmt an, dass bloße Überzeugungen nicht hinreichend und konative Zustände notwendig sind für die Handlungsmotivation. Vertreter rationalistischer Theorien halten dem entgegen, dass auch reine kognitive Zustände (zumindest manchmal) hinreichend sein können zur Handlungsmotivation. In der philosophischen Diskussion wird oftmals der Begriff des Wunsches (*desire*) oder der Pro-Einstellung (*pro-attitude*) als Überbegriff verwendet, der verschiedene Formen von motivationalen Zuständen umfasst. Die Natur von Wünschen (*desires*) wird entsprechend anhand einer motivationalen Theorie erläutert: Wünsche sind dieser gemäß bloß Dispositionen, etwas zu tun. Im Gegensatz zu diesem sehr abstrakten Verständnis von motivationalen Zuständen als Pro-Einstellungen unterschied Aristoteles zumindest zwischen drei großen Bereichen von motivationalen Quellen: körperliche Begierden (*epithymia*) wie Hunger, Durst und sexuelle Bedürfnisse, Emotionen (*thymos*) wie Zorn und Scham sowie Wünsche (*boulesis*), die auf als gut vorgestellte Sachverhalte gerichtet sind. Entscheidend für Willensschwäche ist, dass von diesen drei Quellen nur Wünsche in einer engen Verbindung zu Vorstellungen über das Gute aus Sicht der Person stehen, und in diesem Sinne rational sind – dies schafft Platz für mögliche Konflikte zwischen Wünschen und anderen motivationalen Zuständen: weder sind Wünsche bloß motivationale Zustände, noch sind alle motivationalen Zustände Wünsche.

Dualistische Theorien halte ich für generell besser geeignet als monistische, die Motivation für Handlungen zu erläutern: Es braucht das Zusammenspiel von zumindest zwei distinkten Elementen, kognitive und konative Elemente, damit ein Lebewesen zu einer Handlung oder einem Verhalten motiviert sein kann. Monistische Theorien müssen dem generellen Einwand begegnen, nicht informativ zu sein.<sup>334</sup> Ein Beispiel, das hier lediglich als Strohmann dienen soll: Nach einer Theorie sind Personen dann motiviert, eine Handlung zu tun, wenn sie diese tun *wollen*. Diese Theorie ist monistisch, da sie lediglich mit einem einzigen Element, dem Willen, für die Erklärung der Handlungsmotivation auskommt. In dieser Form ist sie jedoch relativ uninformativ, wenn sie nicht zugleich eine Theorie der Willensbildung anbietet, welche die Genese von einzelnen Willensakten erläutert und verständlich macht. Spätestens hier sollte die

---

<sup>334</sup> Vgl. die Diskussion bei Joseph Heath über das notwendige Ausmaß an Struktur von Entscheidungstheorien. „Decision theory is normally understood to be a conceptual structure used to explain an agent's action in terms of two underlying intentional states: belief and desire. [...] Deliberation is a process through which the agent's set of desires over possible outcomes gets hooked up with a set of beliefs about the possible states, in order to generate a ranking of actions. The beliefs allow the agent to 'project' a set of preferences over outcomes onto the set of available actions, and thus to determine which action is most preferred. It is important to recognize just how conventional this explanatory structure is. Why two intentional states, and not three, or one? For example, there is no reason why we could not develop a decision theory that explained each action, *a*, in terms of a single intentional state, namely, a preference for *a*. Such a theory would be neither interesting nor informative, but there can be no objection to it in principle.” (Heath 2003, 253)

Theorie, um den Mangel an Informativität auszugleichen, das Zusammenspiel verschiedener Elemente für die erfolgreiche Willensbildung annehmen, da ansonsten das Versprechen, eine informative Erklärung zu bieten, bloß auf weitere Ebenen verschoben wird, ohne es je einzulösen. Der Umstand, dass eine solche Theorie den anfänglichen Mangel an Struktur an anderen Orten nachholen muss, ist kein schlagender Einwand gegen solche Theorien, doch liefert er zumindest einen plausiblen Grund für dualistische Theorien: Wenn sowieso mehrere Elemente eingeführt werden, sei es in der Präferenzbildung oder der Bildung von Überzeugungen, dann ist es besser, gleich zu Beginn eine reichhaltigere Struktur anzunehmen.

Timothy Schroeder unterscheidet zwischen sechs verschiedenen Theorien von Wünschen (desires):<sup>335</sup> 1) Handlungsbasierte Theorien (action-based theories): Wünsche sind Dispositionen, etwas zu tun; 2) Lust/Unlust-Theorien (pleasure-based theories): Wünsche sind Dispositionen, unter bestimmten Umständen Lust/Unlust zu empfinden; 3) Wertbasierte Theorien (good-based theories): Wünsche sind identisch mit positiven Werthaltungen (Erscheinungen von etwas als gut); 4) Aufmerksamkeitsbasierte Theorien (attention-based theories): Wünsche sind Dispositionen, die Aufmerksamkeit auf etwas zu lenken, das als gut oder als Handlungsgrund erscheint; 5) Lerntheorien (learning-based theories): Wünsche sind Mechanismen für eine bestimmte Form des Lernens in Organismen; 6) Holistische Theorien (holistic theories): Wünsche sind keine natürlichen Arten, sondern erfüllen je nach Kontext bestimmte Eigenschaften, die von den anderen Theorien für Wünsche hervorgehoben werden.

Schroeder, als Vertreter einer Lerntheorie von Wünschen, kritisiert in *Three faces of desire* (2004) handlungsbasierte Theorien von Wünschen, denen zufolge sich Wünsche von anderen mentalen Zuständen durch zwei Merkmale unterscheiden: 1) Wünsche sind propositionale Einstellungen, sie haben also die Form ‚S wünscht, dass P‘, wobei für P der Ausdruck für eine beliebige Proposition eingesetzt werden kann und S für eine Person steht, und 2) Wünsche sind motivationale Zustände verstanden als Dispositionen, etwas zu tun.<sup>336</sup> Nach 1) sind Wünsche immer auf Sachverhalte gerichtet, die sich durch Propositionen beschreiben lassen. Wünsche können demnach nicht einfach auf Gegenstände (eine Banane) oder Relationen zu Gegenständen (die Geringschätzung meines Nachbarn mir gegenüber) gerichtet sein, sondern nur auf Sachverhalte (‚dass etwas der Fall ist‘). Nach 2) sind Wünsche derart, dass sie sich im Verhalten von Personen zeigen. Schroeder akzeptiert die propositionale Struktur von Wünschen (1), jedoch nicht die Identifikation von Wünschen mit Handlungs- oder Verhaltensdispositionen (2). Gegen letzteres spricht vor allem, dass Personen Wünsche haben können, die sich in keiner Weise in

---

<sup>335</sup> Vgl. Schroeder 2009.

<sup>336</sup> Vgl. Schroeder 2004, 11.

Handlungen äußern.<sup>337</sup> Außerdem ist es auf Basis dieser Theorie schwierig, zwischen Wünschen und anderen motivationalen Zuständen zu unterscheiden. So sprechen zahlreiche Argumente gegen die Reduktion von Intentionen auf Wünsche oder Wunsch-Überzeugungs-Komplexe.<sup>338</sup>

Lust/Unlust-Theorien von Wünschen explizieren Wünsche als Dispositionen, bestimmte Gefühle von Lust/Unlust zu verspüren. ‚S wünscht, dass P‘ heißt, dass S die Disposition hat, Lust (Unlust) zu verspüren, falls P (non-P), bzw. falls sie glaubt, dass P (non-P). Diese Theorie berücksichtigt besser den phänomenalen Charakter von Wünschen, nämlich dass diese meistens mit bestimmten Gefühlen und Affekten verbunden sind, doch gibt es auch hier schwerwiegende Einwände.<sup>339</sup> Der Theorie von Schroeder zur Folge sind Lust/Unlust-Gefühle Repräsentationen der Differenz zwischen erwarteter und tatsächlicher Befriedigung von Wünschen.<sup>340</sup> Wenn Wünsche Objekte von diesen Gefühlen sind, können sie nicht identisch mit Dispositionen sein, diese Gefühle zu empfinden. Oder in einem alltäglichen Verständnis: Lust und Unlust sind kausale *Folgen* der erfolgreichen Erfüllung oder Nichterfüllung von Wünschen, und als solche nicht identisch mit diesen.

Wertbasierte Theorien von Wünschen nehmen an, dass Wünsche identisch sind mit Werturteilen: etwas wünschen heißt es als gut zu bewerten.<sup>341</sup> Auch für Aufmerksamkeitstheorien von Wünschen sind Wünsche stets mit Wertauffassungen des gewünschten Objekts verbunden: etwas wünschen heißt hier, die Disposition zu haben, an das gewünschte Objekt zu denken und es als Grund für weitere Handlungen anzusehen.<sup>342</sup> Für beide Theorien stellen willensschwache Handlungen besondere Herausforderungen dar. Ein weiteres Problem ist die Anwendung des Wunschbegriffes auf nicht-menschliche Lebewesen. Diese haben Wünsche, jedoch ohne über die geeigneten kognitiven Kapazitäten für entsprechende Werturteile zu verfügen.<sup>343</sup>

---

<sup>337</sup> Vgl. Schroeder 2004, 16ff. Schroeder bringt insgesamt sieben Einwände gegen die Theorie. Vgl. auch Schroeder 2004, 15–27.

<sup>338</sup> Vgl. Bratman 1987, ch. 2. Für eine Verteidigung einer solchen Reduktion vgl. Ridge 1998.

<sup>339</sup> Vgl. Schroeder 2004, 27–35.

<sup>340</sup> Vgl. Schroeder 2004, ch. 3.

<sup>341</sup> Vgl. Kap. III.5.3; Tenenbaum 2007.

<sup>342</sup> Vgl. Scanlon 1998, ch. I.1.8. „What I am claiming, however, is not that all desires arise from prior judgments but rather that having what is generally called a desire involves having a tendency to see something as a reason.” (Scanlon 1998, 39) Scanlon unterscheidet zwischen Wünschen in ihrer Rolle für die Aufmerksamkeitslenkung (directed-attention sense) von Wünschen in einem bloß formalen Sinn als Pro-Einstellung bzw. gemäß einer handlungsbasierten Theorie von Wünschen.

<sup>343</sup> „One puzzle for such theories might be to explain the relationship of desires to non-human animals. On the one hand, it would seem that rats desire to get away from cats, desire to be around other rats, and the like. On the other hand, it would seem that rats do not represent anything as good (they would both seem to lack the concept of goodness and to lack a perceptual-style representation of goodness that would be well poised to generate such a concept). But if rats can desire without representing the good, then why would people be different?” (Schroeder 2009)

Nach Schroeder zeichnen sich Wünsche vor allem anhand von drei Merkmalen gegenüber anderen mentalen Zuständen aus: ihre motivationale Kraft, ihre Verbindung zu Gefühlen der Lust/Unlust und ihre Rolle für Mechanismen der Belohnung und Strafe.<sup>344</sup> Anstatt sich auf die Rolle von Wünschen für die Handlungsmotivation sowie für das Fühlen von Lust und Unlust zu konzentrieren, entwickelt er seine Theorie von Wünschen anhand deren Rolle für Prozesse von Belohnung und Strafe aufgefasst als eine spezifischen Art von Lernen. Etwas zu wünschen (verabscheuen) heißt nach dieser Theorie, es als Belohnung (Strafe) aufzufassen.<sup>345</sup> Etwas als Belohnung (Strafe) aufzufassen heißt wiederum, eine spezifische Form des Lernens gegenüber dem Objekt zu zeigen. Diese Theorie inkorporiert die motivationalen und phänomenalen Aspekte der anderen Theorien, ohne deren hauptsächliche Schwächen zu teilen. Da diese Form des Lernens oft mit Verhalten bzw. Handlungen verknüpft ist, sind Wünsche auch motivational bedeutsam. Da Wünsche aber nicht als Handlungsdispositionen aufgefasst werden, können Personen auch Wünsche gegenüber Objekten hegen, die sich niemals in entsprechendem Verhalten ausdrücken. Da das Erhalten von Belohnungen oftmals mit Lust einhergeht, sind Wünsche eng mit Lust/Unlust-Empfindungen verknüpft, aber erstens sind diese Gefühle von den entsprechenden Wünschen verschieden und zweitens kann die generelle Empfindsamkeit von Personen unabhängig von ihrem Set an Wünschen variieren, wie z.B. in Fällen von Depression.

Welche Relevanz haben diese verschiedenen Theorien von Wünschen für das Verständnis von Willensschwäche? Handlungsbasierte Theorien, Lust/Unlust-Theorien und Lerntheorien können die Motivation für willensschwache Handlungen erläutern, da sie keinen strikten Zusammenhang zwischen der kognitiven Einschätzung einer Person für eine Handlung und der Motivation, diese zu tun, annehmen. Angesichts der anderen Probleme der handlungsbasierten Theorien und der Lust/Unlust-Theorien stellt die Lerntheorie eine attraktive Variante dar. Werttheorien von Wünschen stehen in Konflikt zu Willensschwäche durch die Annahme einer strikten Korrelation zwischen Wertnehmung und Wünschen.

Neben Wünschen sind *Emotionen* und *Intentionen* weitere konative Quellen der Handlungsmotivation. Emotionen haben eine motivationale Komponente, da sie je nach spezifischem Typ von Emotion mehr oder weniger stark durch Handlungen ausgedrückt werden bzw. mit Handlungsmustern in Beziehung stehen. Dies kann oftmals indirekt über das Auslösen eines entsprechenden Wunsches geschehen, aber es sind auch direkte Formen der Handlungs-

---

<sup>344</sup> Vgl. Schroeder 2004, 35.

<sup>345</sup> Die Definition der 'Reward Theory of Desire': „To have an intrinsic (positive) desire that *P* is to use the capacity to perceptually or cognitively represent that *P* to constitute *P* as a reward. To be averse to it being the case that *P* is to use the capacity to perceptually or cognitively represent that *P* to constitute *P* as a punishment.“ (Schroeder 2004, 131)

motivation plausibel.<sup>346</sup> Emotionen erfüllen neben dieser motivationalen Funktion vor allem auch informative Funktionen und sozial-kommunikative Funktionen.<sup>347</sup> *Affekte* sind gemeinhin Emotionen, die besonders intensiv sind und kurzfristig auftreten. Auch diese stehen natürlich in enger Beziehung zu spezifischen Handlungen und Verhaltensmustern. Intentionen sind unbestritten motivationale Zustände, da sie sowohl in Form von langfristigen Plänen wie auch als handlungsleitende Absichten direkt mit Handlungen verknüpft sind. Sie zeigen dabei Eigenschaften, die sie von Wünschen unterscheiden.<sup>348</sup> So können einander logisch widersprechende Sachverhalte nicht intendiert werden, wenn die Person über ihre Widersprüchlichkeit Kenntnis hat, während dies bei Wünschen nicht der Fall ist. So kann z.B. Markus wünschen, dass es heute regnet, und auch, dass es nicht regnet. Auch dürften Intentionen zumindest schwach aggregierend sein: Wenn eine Person eine Handlung A intendiert und ebenso eine Handlung B, dann intendiert sie normalerweise die Kombination der beiden Handlungen (A & B).<sup>349</sup> Auch dieses Prinzip ist für Wünsche nicht gültig: Peter kann wünschen, Fußball zu spielen und ebenso wünschen, Sartre zu lesen, und dennoch nicht wünschen, Fußball zu spielen und Sartre zu lesen. Von Intentionen lassen sich *Strebungen* unterscheiden: dies sind Wünsche, die sich bereits auf spezifische Handlungsmuster richten, aber noch nicht den speziellen Bedingungen von Intentionen genügen. Trotz der speziellen kognitiven Ansprüche, die Intentionen erfüllen, sind sie dennoch keine kognitiven Zustände wie z.B. spezielle Glaubenszustände.

Diese Liste an motivationalen und für die Motivation wichtigen mentalen Zuständen zeigt, dass von einer Pluralität von Quellen der Motivation ausgegangen werden sollte und sich diese nicht immer auf ein bestehendes evaluatives Urteil der Person zurückverfolgen lassen. Der motivationale Externalismus kann dabei sowohl synchrone motivationale Willensschwäche integrieren, als auch den Normalfall des Handelns auf Basis von evaluativen Urteilen erläutern.

## **Zusammenfassung**

Selbsttäuschungstheorien von Willensschwäche interpretieren Willensschwäche als Täuschung von Personen über sich selbst: die willensschwache Person handle nicht wider ihr besseres Wissen, verstanden als situativ bestes Urteil über ihre Handlungsoptionen, sondern *gemäß* diesem Urteil. Die willensschwache Person verbirgt sich aber diesen Umstand vor sich selbst, indem sie vorher und/oder nach erfolgter Handlung ihren Meinungsbildungsprozess irrational beeinflusst. Die Beeinflussung ist dabei motiviert durch den Wunsch, ein bestimmtes Selbstbild von sich

---

<sup>346</sup> Dies gilt für einen engen Begriff von Wunsch, nicht für Pro-Einstellung, da letztere so definiert sind, dass stets eine Pro-Einstellung vorliegt, wenn jemand handelt.

<sup>347</sup> Vgl. Rothermund & Eder 2011, 175–180.

<sup>348</sup> Vgl. die Diskussion bei Bratman 1987, ch. 2. Aber: Velleman 1985.

<sup>349</sup> Vgl. jedoch Bratman 1987, ch. 8.

aufrechterhalten zu können, welches mit der willensschwachen Handlung in Konflikt steht. Während sich nach Wolf die Person über ihre tatsächlichen Präferenzen täuscht, ist nach Schälike die Selbstzuschreibung von Willensschwäche selbst Teil einer Selbsttäuschungsstrategie. Diese erlaube theoretisch eine Kluft zwischen evaluativen Bewertungen und Handlungsmotiven, die realiter gar nicht bestehe, und damit Raum für selbstgenügsame Fehlinterpretationen. Selbsttäuschungstheorien verändern damit die Beschreibung des Phänomens der Willensschwäche zugunsten einer alternativen Beschreibung, die kompatibel ist mit monistischen Handlungstheorien. Innerhalb solcher Theorien werden Urteile von Personen über ihre situativ beste Handlungsoption als *Intentionen*, diese Handlung zu vollziehen, verstanden. Intentionen werden wiederum so verstanden, dass sie vollzogen werden, wenn die Person entsprechend handeln kann. Daraus ergibt sich, dass Personen stets das tun, was sie situativ als beste Handlungsoption beurteilen, wenn sie entsprechend handeln können. Selbsttäuschungstheorien sind auch außerhalb monistischer Handlungstheorien interessant, da nicht auszuschließen ist, dass Willensschwäche tatsächlich mit Selbsttäuschung gemeinsam auftreten kann bzw. dass in einigen Fällen von Willensschwäche die Person sich täuscht, wider ihr besseres Wissen gehandelt zu haben. Der Anspruch, *alle* Fälle von Willensschwäche auf diese Weise explizieren und erklären zu wollen, hängt jedoch von diesen Selbsttäuschungstheorien unabhängigen handlungstheoretischen und metaethischen Prämissen ab. Wer diese nicht teilt, der schafft sich zumindest vorerst Spielraum, Willensschwäche als Handeln wider das aktuell Beste zu verstehen.

Platons Sokrates argumentiert auf Basis eines psychologischen Hedonismus und einer Identifikation von Wertprädikaten mit Prädikaten über Lust- bzw. Unlustzuständen gegen die Möglichkeit synchroner Willensschwäche. Da die Erkenntnis des Guten bereits hinreichend sei für deren Ausführung, kann eine Person nicht wider besseres Wissen handeln. Was willensschwachen Personen dagegen fehle, ist die Messkunst, zukünftige Lust und Unlust gegen momentane Lust und Unlust richtig abzuwägen. Denselben Gedanken in moderner ökonomischer und entscheidungstheoretischer Terminologie greift George Ainslie auf. Seine These ist, dass Willensschwäche aus einer systematisch irrationalen Weise des Diskontierens von zukünftigem gegenüber gegenwärtigem Nutzen resultiert. Anstatt exponentiell diskontieren Personen und andere Lebewesen in einer Weise, die sich besser durch eine hyperbolische Kurve beschreiben lässt, welche die Eigenschaft hat, zeitlich nahe Gewinne unverhältnismäßig höher zu bewerten als ferne, was wiederum zu irrationalem Präferenzwechsel in Form von Willensschwäche führt. Während bei Platon Evaluation und Motivation eine Einheit bilden, die jeden echten motivationalen Konflikt verunmöglicht, ist bei Ainslie die Person lediglich eine Ansammlung an Präferenzen, die miteinander strategisch interagieren. Diese metaphorische Teilung einer Person in verschiedene Sub-Personen führt dazu, dass synchrone Konflikte nicht

mehr modellierbar sind innerhalb dieser Theorie, da die dafür notwendiger Einheit einer Person nicht in diesem Modell repräsentierbar ist. Dennoch leisten Diskontierungstheorien wichtige Lösungsansätze für das Verständnis diachroner motivationaler Konflikte und für viele Phänomene, die mit dem plötzlichen Umschwung von Wertungen, der Änderung des besseren Urteils, einhergehen.

Nach Richard Hare folgt aus der Präskriptivität moralischer Urteile, dass Willensschwäche als Handeln wider Moralurteile, welchen die Person aufrichtig und im vollen Sinn zustimmt, unmöglich ist. Moralurteile sind universelle Präskriptionen des eigenen wie des Verhaltens anderer, die in hinreichend ähnlichen Situation Antworten auf die Frage geben, wie gehandelt werden soll. Die Präskriptivität lässt sich dabei verstehen in Analogie zum vorschreibenden Charakter von Imperativen. Imperative, die an sich selbst gerichtet sind, werden, so die Argumentation, aber auch umgesetzt, wenn die betreffende Person physisch und psychisch dazu in der Lage ist. Robert Dunn kritisiert das Konzept in drei Punkten: *Erstens* seien evaluative Urteile überhaupt nicht mit direktiven, für Imperative typischen Illokutionen notwendig verbunden. *Zweitens* sind direktive, für Imperative typische Illokutionen ungeeignet, um Intentionen auszudrücken, die das *eigene* Handeln betreffen. *Drittens* sind evaluative Ausdrücke generell hauptsächlich deskriptiv zu verstehen, und daher semantisch nicht notwendigerweise mit kommissiven, direktiven, expressiven oder konativen Illokutionen verbunden, d.h. sie drücken neben ihrer beschreibenden Funktion nicht notwendigerweise Intentionen, Empfehlungen, Befehle, Wünsche usw. des Sprechers an sich oder andere aus.

Inkommensurabilitätsthesen erläutern die Möglichkeit von Willensschwäche durch den Hinweis auf die Existenz inkommensurabler Werte und den damit verbundenen Gütern. Inkommensurabilität bedeutet dabei das Fehlen eines einheitlichen (skalaren) Maßstabes, mit dessen Hilfe der Abstand zwischen beliebigen Paaren an Werten anhand eines einheitlichen Maßes gemessen werden kann. So verstandene Inkommensurabilität impliziert nicht die Inkomparabilität derselben Wertpaare. Inkommensurabilität mache Willensschwäche deshalb möglich, da die Handlungsoptionen, zwischen denen die willensschwache Person wählt, zwar vergleichbare, aber inkommensurable Wertrealisierungen einschließen. Das heißt, dass die willensschwache Handlung Wertrealisierungen verspricht, die *nicht* quantitativ oder qualitativ in der besser bewerteten Handlungsoption enthalten sind. Da Werte zudem Gründe für Handlungen liefern, habe die willensschwache Person daher einen Grund für ihre Handlung, wenn auch einen schlechten.

Gegen diese Art von Theorien kann eingewendet werden, dass Inkommensurabilität von Werten weder hinreichend noch notwendig ist, um die Möglichkeit von Willensschwäche zu erklären.

*Erstens:* Ohne die Annahme der Fragmentierung des Psychischen bzw. einer Unterscheidung in kognitive und affektive Dimensionen der Bezugnahme auf die verschiedenen Handlungsoptionen ist unklar, wie die willensschwache Person motiviert sein könnte für die willensschwache Handlung. Wenn Motivation und Evaluation einheitlich miteinander korreliert wären, dann müsste die besser bewertete Option auch attraktiver sein aus Sicht der handelnden Person. Es besteht demnach ein Dilemma: sind die Handlungsoptionen vergleichbar (Komparabilität), dann würde die Einheitsthese von Motivation und Evaluation Willensschwäche verunmöglichen. Sind die Handlungsoptionen nicht vergleichbar, gäbe es gar kein Problem mit willensschwachen Handlungen, da generell keine Handlungsoption als ‚besser als‘ die jeweils anderen ausgezeichnet werden könnte. *Zweitens:* Bereits die These der Fragmentierung des Psychischen reicht aus, um die Möglichkeit von Willensschwäche zu erläutern, auch unter der Annahme vollständiger Kommensurabilität von Werten. So könnte selbst ein vollständiger Hedonist die Handlungsoption, die weniger Lust verspricht, attraktiver finden, wenn affektive und kognitive Bezugnahmen auf die jeweiligen Optionen auseinandergehen können. Damit sind Inkommensurabilitätsthesen nicht geeignet, die Möglichkeit von Willensschwäche zu erläutern, wenn sie auch sonst plausibel sein mögen als generelle Thesen über die Pluralität von Werten, Gütern und deren gegenseitige Vergleichbarkeit und Messbarkeit.

Für Urteils-Internalisten besteht ein begrifflich notwendiger Zusammenhang zwischen evaluativem Urteil und der Motivation, entsprechend zu handeln. Dieser Zusammenhang kann verschieden stark interpretiert werden. Da Internalisten nur auf Eigenschaften des evaluativen Urteils zur Erläuterung des internen Zusammenhangs von Evaluation und Motivation zurückgreifen können, formulieren sie den Zusammenhang entweder proportional oder nicht-proportional. Für die proportionale Variante ist Willensschwäche ein direktes Gegenbeispiel, die nicht-proportionale Variante kann dagegen die handlungsleitende Funktion von evaluativen Urteilen nicht adäquat erläutern. Es bleibt, die internalistische These selbst abzuschwächen: Schwache Internalismen machen den Zusammenhang zwischen evaluativem Urteil und der Handlungsmotivation von der Erfüllung weiterer Bedingungen – die Person ist praktisch rational, psychisch stabil etc. – abhängig. Das Problem ist, dass diese Bedingungen bloße ad hoc Modifikationen oder Trivialisierungen der ursprünglichen These darstellen und damit nicht geeignet sind, die internalistische These vor dem Einwand auf Basis von Willensschwäche zu bewahren. Motivationale Urteils-Externalisten behaupten daher, dass der Zusammenhang zwischen Evaluation und Motivation durch einen zusätzlichen, dem evaluativen Urteil externen Faktor gestiftet wird. Dieser Faktor ist eine Pro-Einstellung, die als Wunsch interpretiert werden kann. Gemäß einer Lerntheorie von Wünschen sind Wünsche mentale Zustände, welche Sachverhalte als Belohnungen auffassen und mit einer spezifischen Art des Lernens verknüpft

sind. Zudem sind diese oftmals motivational wirksam und verbunden mit Gefühlen von Lust bzw. Unlust. Ein externalistischer Ansatz ist daher von allen untersuchten Ansätzen am besten geeignet, um die Möglichkeit von synchroner motivationaler Willensschwäche zu erläutern und in ein Modell des Verhältnisses von Evaluation und Motivation einzubetten.

## IV Problemfeld Intentionalität

Das Phänomen der Willensschwäche wird nicht nur als Testfall für Theorien über das Verhältnis von Evaluation und Motivation angesehen, sondern auch als handlungstheoretischer Testfall für das rechte Verständnis der Intentionalität von Handlungen.

„The *akrates* acts intentionally. So we will not have a good theory of *akrasia* until we have a good theory of intention; and conversely the requirement that it leave room for the possibility of *akrasia* is a condition of adequacy on an account of intention.” (Peacocke 1985, 51)

Die Gewichtung dieser Korrelation ist Teil der Debatte, in der eine Seite handlungstheoretische Prämissen als Restriktion für mögliche Formen von Willensschwäche benutzt, während für die andere Seite mögliche Formen von Willensschwäche als Grenzfälle für die Adäquatheit handlungstheoretischer Prämissen fungieren. In jedem Fall wirft Willensschwäche grundlegende Fragen für Handlungstheorien auf: Ist es, und wenn ja, wie ist es möglich, dass eine Person *intentional* wider ihr eigenes besseres oder bestes Urteil bzw. Wissen handelt? Was bedeutet hier ‚intentional handeln‘? Handelt die willensschwache Person dabei aus einem Grund, und wenn ja, aus welchem? Diese und weitere Fragen bilden das Problemfeld der Intentionalität innerhalb der Diskussion der Willensschwäche. Dieses Problemfeld ist dabei verschieden von der Frage nach der Verantwortung für Willensschwäche und willensschwache Handlungen. Obwohl intentionale Handlungen paradigmatische Objekte von Verantwortungszuschreibungen sind, füllen sie nicht die gesamte Klasse an diesen Objekten aus. Ich halte demnach den Bereich, für den Personen verantwortlich sind, für wesentlich größer als die Gesamtheit intentionaler Handlungen. Und selbst wenn die einzigen Gegenstände, für die wir verantwortlich sein können, intentionale Handlungen sind, so sind wir nicht für jede unserer intentionalen Handlungen verantwortlich. Die Existenz von Entschuldigungen implizieren gerade, dass wir nur für einen Teil unserer Handlungen die Verantwortung übernehmen. Die Absichtlichkeit von Handlungen ist demnach nicht hinreichend für die Zuschreibung von Verantwortung.<sup>350</sup>

Zuerst (1) formuliere ich anhand der Unterscheidung von Elizabeth Anscombe zwischen verschiedenen Gebrauchsweisen des Begriffs der ‚Intention‘ in Bezug zu Handlungen wesentliche Fragestellungen für die Intentionalität willensschwacher Handlungen. In (2) untersuche ich die handlungsleitende Intention, mit der willensschwache Personen ihre Handlungen vollziehen. Dabei diskutiere ich verschiedene skeptische Positionen, welche gegen diese Art der Intentionalität willensschwacher Handlungen argumentieren. In (3) unternehme ich einen kleinen Exkurs hinsichtlich der Frage, was praktische Gründe eigentlich sind. In (4) diskutiere ich Fragen der Charakterisierung willensschwacher Handlungen als absichtlich. In (5) untersuche ich die

---

<sup>350</sup> Vgl. Kap. V.

Rolle von Intentionen, verstanden als Pläne für zukünftige Handlungen, für das Phänomen der Willensschwäche In (6) widme ich mich dem Thema exekutiver Willensschwäche und diskutiere entsprechende Ansätze, vor allem jenen von Richard Holton.

## IV.1 Handlungstheoretische Grundlagen

Elizabeth Anscombe hat mit ihrem Buch *Intention* die handlungstheoretische Diskussion im letzten Jahrhundert entscheidend geprägt. Grundsätzlich unterscheidet sie darin zwischen drei Weisen, in denen der Begriff der Intention gebraucht wird in Bezug zu Handlungen:

„Very often, when a man says ‘I am going to do such-and-such’, we should say that this was an expression of intention. We also sometimes speak of an action as intentional, and we may also ask with what intention the thing was done. In each case we employ a concept of ‘intention’; now if we set out to describe this concept, and took only one of these three kinds of statement as containing our whole topic, we might very likely say things about what ‘intention’ means which it would be false to say in one of the other cases.“  
(Anscombe 1963, 1.)

Anscombe unterscheidet in Folge in Bezug auf ‚intentionale Handlungen‘ (intentional actions) zwischen:

- 1) Intentionen als Pläne:<sup>351</sup> Intentionen als Absichten, etwas in (naher) Zukunft zu tun. Beispiele: ‚Ich gehe morgen ins Kino‘; ‚Ich fliege in einer Woche nach Spanien‘; ‚Nächstes Jahr kaufe ich mir ein neues Auto‘.
- 2) Intentionen als handlungsleitende Absichten:<sup>352</sup> Intentionen als Absichten, hier und jetzt etwas zu tun, mit denen jemand etwas tut, zu welchem Zweck jemand eine Handlung ausführt oder aus welchem Grund jemand etwas tut.<sup>353</sup> Beispiele: ‚Ich gehe gerade ins Kino‘; ‚Ich gehe ins Konzert um mich zu unterhalten‘; ‚Ich starte den Computer, weil ich ein Email schreiben will‘.<sup>354</sup>
- 3) ‚Intentional‘ als Eigenschaft von Handlungen oder als Relation zwischen Handlungsbeschreibungen und Einstellungsbeschreibungen:<sup>355</sup> Die Zuschreibung von Absichtlichkeit, mithin als Charakterisierung von Verhaltensweisen als absichtlich oder unabsichtlich. Beispiele: ‚Du bist mir absichtlich auf den Fuß gestiegen‘; ‚Ich habe das nicht absichtlich getan!‘

---

<sup>351</sup> ‘Prior intentions’ (Searle 1983); ‘distal intentions’ (Mele 2010b); ‘pure intendings’ (Davidson 1978).

<sup>352</sup> ‘Intention in action’ (Searle 1983); ‘proximal intention’ (Mele 2010b).

<sup>353</sup> Diese Gleichsetzung von Zweck und Grund einer Handlung ist auf den ersten Blick nicht verständlich. Aber für die kausale Handlungstheorie sind Zwecke in den Ursachen von Handlungen in Form von Wünschen und anderen mentalen Zuständen repräsentiert.

<sup>354</sup> Auch handlungsleitende Absichten haben die Struktur von Plänen. Ich werde jedoch nur Intentionen bzw. Absichten des Typs 1 als ‚Pläne‘ bezeichnen, um terminologisch eindeutig zwischen den beiden Arten von Intentionen zu unterscheiden. Auch Pläne, also auf zukünftige Handlungen gerichtete Absichten, spielen eine Rolle für handlungsleitende Absichten, und sind auch meist präsent während eine Person handelt.

<sup>355</sup> ‚Action as intentional‘. Einstellungsberichte wären Berichte über mentale Zustände der Person, welche die entsprechenden Handlungen rationalisieren. Vgl. Löhner 2003, 31–58. Die Handlung an sich ist nicht intentional, sondern nur in Relation zu diesen Einstellungsbeschreibungen.

Anscombe geht es dabei darum, das Verhältnis dieser verschiedenen Gebrauchsweisen zueinander zu untersuchen, wobei sie davon ausgeht, dass diese nicht einfach bloße Äquivokationen darstellen, sondern systematisch miteinander zusammenhängen. Eine Handlung ist nach Anscombe nur dann intentional, wenn über sie sinnvoll die Frage gestellt werden kann, weshalb sie vollzogen wurde.

„What distinguishes actions which are intentional from those which are not? The answer that I shall suggest is that they are the actions to which a certain sense of the question ‘Why?’ is given application; the sense is of course that in which the answer, if positive, gives a reason for acting. But this is not a sufficient statement, because the question “What is the relevant sense of the question ‘Why?’” and “What is meant by ‘reason for acting?’” are one and the same.” (Anscombe 1963, 9)

Ob etwas eine Handlung ist oder nicht, ergibt sich daraus, ob eine bestimmte Art von ‘Warum-Frage’ angemessen ist oder nicht. Diese wird wiederum beantwortet durch die Angabe von Gründen der Person für ihr Handeln. Intentionales Handeln ist demnach Handeln aus Gründen. Um intentionales Handeln näher bestimmen zu können, müssen die Gründe, aus denen es erfolgt, näher charakterisiert werden. Innerhalb kausaler Handlungstheorien sind dies mentale Ereignisse, nämlich Überzeugungen und Wünsche, bzw. das Eintreten von diesen, welche Gründe für die jeweiligen Handlungen sind, wenn sie diese verursachen. Innerhalb teleologischer Handlungstheorien sind dies Ziele, auf die das Handeln gerichtet ist und die es erklären und rationalisieren. Die Explikation des Handlungsbegriffs über den Begriff des Handlungsgrundes ist natürlich nicht konkurrenzlos. So ist philosophisch umstritten, ob der Begriff der handlungsleitenden Absicht auf den des Handlungsgrundes bzw. des Handelns aus Gründen reduziert werden kann oder nicht. Hier sollte diese kurze Einführung genügen, um über das Problemfeld anhand einer systematischen Fragestellung, der nach der verschiedenen Bedeutung von Intentionalität in Bezug zu Handlungen, einen Überblick zu gewinnen.

Im Anschluss an die Differenzierung verschiedener Verwendungsweisen von ‚Intentionalität‘ möchte ich für die Frage nach der Intentionalität willensschwacher Handlungen zumindest drei Fragen voneinander unterscheiden, die ich der Reihe nach untersuchen möchte:

- 1) Verfolgt die Person mit ihrer willensschwachen Handlung eine bestimmte handlungsleitende Absicht? Wenn ja, was ist diese? Handelt die willensschwache Person aus Gründen?
- 2) Vollzieht die Person die willensschwache Handlung absichtlich? Wenn ja, worin gründet die Zuschreibung von Absichtlichkeit? Unter welcher Beschreibung sind willensschwache Handlungen intentional?
- 3) Formt die Person vor ihrer willensschwachen Handlung eine Intention im Sinne eines Planes, diese Handlung zu tun? Wäre dies überhaupt möglich? Wenn ja, was ist der Inhalt dieses Planes?

Diese drei Fragen sind verschieden und zielen letztlich auf etwas Verschiedenes ab. Dies ist nicht trivial, da viele handlungstheoretische Diskussionen die Frage betreffen, ob diese drei Momente nicht zusammenfallen bzw. nicht eines oder mehrere dieser drei Momente durch eines der anderen expliziert werden könnte. Vorausgesetzt jemand akzeptiert, dass es Sinn macht, von Intentionen (Plänen) unabhängig von Handlungen zu sprechen, also von Intentionen als in irgendeiner Weise von konkreten Handlungen unabhängige Entitäten, so meine ich, dass diese drei Elemente auf Verschiedenes verweisen. Dies lässt sich zeigen, indem keines der drei Elemente hinreichend für die Zuschreibung der jeweils anderen ist. Folgende Beziehungen bestehen meiner Meinung nach zwischen den verschiedenen Elementen:

a) Jemand kann einen Plan haben, z.B. am nächsten Tag eine Bank auszurauben, ohne diesen Plan zu verwirklichen, also die Bank tatsächlich auszurauben. D.h. jemand hat einen Plan, ohne deshalb eine entsprechende handlungsleitende Absicht (intention in action) zu bilden, noch eine entsprechende Handlung absichtlich zu tun.

b) Jemand kann ohne einen Plan intentional handeln, d.h. mit einer handlungsleitenden Absicht, die nicht einem vorhergehenden Plan entspricht. Nicht alle Handlungen gehen auf Überlegungen zurück oder auf das Fassen eines Planes, sondern entstehen spontan oder aus Gewohnheit. Also ist die Intention, mit der jemand handelt, nicht notwendigerweise identisch mit einem der Handlung vorhergehenden Plan, diese Handlung zu tun. Die Handlung ist dabei nur absichtlich unter der Beschreibung, auf welche die handlungsleitende Intention gerichtet ist.

c) Jemand kann einen Plan fassen, diesen Plan umsetzen, und dennoch wird die Handlung als unabsichtlich charakterisiert. Das Vorliegen einer handlungsleitenden Intention mag eine notwendige Bedingung für die Zuschreibung von Absichtlichkeit sein, sie ist jedoch nicht hinreichend: bei jeder absichtlichen Handlung intendiert die Person etwas im Handeln, doch umgekehrt ist nicht jede Handlung, die mit einer Absicht vollzogen wurde, auch absichtlich. Dies deshalb, da Handlungen über ihr spezifisches Handlungsergebnis beschrieben und individuiert werden, und das Eintreten von diesem Resultat von kausalen Faktoren abhängt, die nur bedingt durch den Akteur beherrscht werden können. Die Entscheidung dieser Frage hängt davon ab, wie das Verhältnis von Handlungen und Handlungsfolgen genauer bewertet wird. Dies zeigt das Problem abweichender Kausalketten:<sup>356</sup> in der Charakterisierung von Handlungen als absichtlich oder unabsichtlich ist ein normatives Element enthalten, das Bezug auf praktisches Wissen und die Umwelt nimmt, in der eine Handlung stattfindet. Hierbei spielen Normalitätsvorstellungen eine große Rolle, also was bestimmte Arten von Akteuren normalerweise zu leisten imstande sind und was nicht. Diese Vorstellungen sind letztlich relativ zu kulturellen und sozialen

---

<sup>356</sup> Vgl. Kap. II.2.4.

Rahmenverhältnissen. Dies heißt, dass die Frage, ob eine Handlung absichtlich vollzogen wurde, nicht alleine auf der Basis individueller mentaler Zustände entschieden werden kann: eine Person kann das, was sie tut, nicht von selbst zu etwas Absichtlichem machen, sondern nur etwas tun, was normalerweise innerhalb einer bestimmten Gruppe als absichtlich charakterisiert werden kann. Dies ist ernüchternd für kausale und teleologische Handlungstheorien, welche notwendige und hinreichende Bedingungen für das Stattfinden von absichtlichen Handlungen formulieren möchten, und dieses Ziel als erreicht ansehen, wenn sie die Intention, mit der eine Person handelt – was die Person mit ihrer Handlung intendiert – explizieren. Eine Analogie mag dabei helfen, diesen Gedanken näher zu verstehen. Eine Person könnte die Absicht haben, eine Gruppe von Leuten zum Lachen zu bringen, indem sie einen Witz erzählt. Die Person tut dies, doch lachen die Leute nicht deswegen, weil sie den Witz selbst komisch finden, sondern weil sie die Art und Weise der Person, den Witz zu erzählen, witzig finden. Das Erzählen des Witzes verursacht kausal das intendierte Resultat – die Leute lachen –, jedoch nicht auf eine ‚normale‘ Weise. Die Person hat, obwohl sie die Absicht hatte und entsprechend dieser Absicht etwas intentional getan hat – den Witz zu erzählen – nicht absichtlich gehandelt hinsichtlich der Beschreibung, unter der die Handlung intendiert war – die Leute zum Lachen zu bringen *durch* das Erzählen eines Witzes –, *obwohl* sie das intendierte Resultat hervorbrachte.

Zusammenfassend vertrete demnach die Ansicht, dass folgende Abhängigkeiten zwischen den verschiedenen Verwendungsweisen von ‚Intention‘ und ‚intentional‘ für Handlungen gelten:

- 1) Auf die Zukunft gerichtete Handlungspläne sind weder notwendig noch hinreichend sowohl für intentionale Handlungen als auch für die Charakterisierung von Handlungen als absichtlich.
- 2) Handlungsintentionen sind nicht hinreichend, jedoch notwendig für die Charakterisierung von Handlungen als absichtlich. Dieser Befund ergibt sich aus dem Problem abweichender Kausalketten und ist unabhängig davon, ob eine kausale oder teleologische Handlungstheorie verfolgt wird.
- 3) Die Charakterisierung einer Handlung als absichtlich ist nicht hinreichend für die Zuschreibung eines Handlungsplanes, jedoch hinreichend für die Zuschreibung einer handlungsleitenden Intention. Bei jeder absichtlichen Handlung gibt es etwas, was die Person absichtlich tut, d.h. eine handlungsleitende Intention.<sup>357</sup> Letztere ist die Beschreibung der Handlung aus Sicht der Person.

---

<sup>357</sup> Vgl. Richard Holton über den Unterschied zwischen ‚Intention‘ und ‚intentional‘: „Nonetheless, there are very good grounds here for thinking that there are two different notions in play: one, normally denoted by the noun ‚intention‘, that is a psychological notion; and one, normally denoted by the adverb ‚intentionally‘, that combines the psychological with the normative in some rather complicated way.” (Holton 2009, 16)

## IV.2 Handlungsleitende Intention und Willensschwäche

Arthur F. Walker hat in einem Überblicksartikel *The problem of weakness of will* (1989) hilfreich zwischen verschiedenen skeptischen Ansätzen unterschieden, welche die Intentionalität willensschwacher Handlungen anzweifeln.<sup>358</sup> Deren Argumente basieren auf 1) einer ‚backward connection‘ bzw. Rückwärtsverbindung zwischen intentionalen Handlungen und besseren Urteilen;<sup>359</sup> 2) einer ‚forward connection‘ bzw. Vorwärtsverbindung zwischen besseren Urteilen und intentionalen Handlungen; 3) einer ‚reason theory of intentional action‘ bzw. Gründe-Theorie intentionalen Handelns. Ich verstehe diese verschiedenen Zweifel hier so, dass sie in erster Linie in Frage stellen, dass die Person eine konkrete Absicht haben kann, die ihre willensschwache Handlung leitet (intention in action). Dies sollte nicht gleichgesetzt werden mit der Frage, ob nun die Person absichtlich gehandelt hat oder nicht, auch wenn diese beiden Momente von verschiedenen Autoren vermischt werden. Fraglich ist also, ob die Person überhaupt etwas absichtlich tun kann, wenn sie willensschwach handelt, ob sie demnach das, was sie tut, intendieren kann oder nicht.

### IV.2.1 Backward connection

Theorien auf Basis der backward connection behaupten einen strikten Zusammenhang zwischen dem Vollzug einer intentionalen Handlung und dem Umstand, diese Handlung als beste oder als besser zu beurteilen. Bestimmte Werturteile seien demnach notwendige Bedingungen für intentionale Handlungen. Das Konditional lässt sich so formulieren: Wenn eine Person intentional (und ungezwungen) handelt, dann beurteilt sie die Handlung auch als beste bzw. als besser gegenüber allen anderen.

„The Backward Connection argument runs as follows:

- (1) If *S* intentionally *As*, then *S* judges that *A*ing is best (at the time he acts). (This is the backward connection thesis.)
- (2) If *S* intentionally *As* against his better judgment (i.e., *As* akratically), then *S* judges that *A*ing is not best (at the time he acts).
- (3) It is (nearly) impossible that *S* both judges that *A*ing is best and judges that *A*ing is not best at the same time.
- (4) Therefore, last ditch akratic action is impossible.“ (Walker 1989, 658)

Prämisse 3 ist deshalb interessant, da der Widerspruch, der dem Akteur anhand der backward connection zugewiesen wird, das gleichzeitige Fällen von einander widerstreitenden Urteilen ist.

---

<sup>358</sup> Walker nennt auch Theorien, welche die Freiheit oder Verantwortlichkeit der willensschwachen Handlung anzweifeln. Dies untersuche ich in Kap. V.

<sup>359</sup> Walker führt noch eine ‚inference theory of intentional action‘, eine Inferenz-Theorie intentionalen Handelns, an, fügt jedoch hinzu, dass das Argument auf Basis einer solchen Theorie – intentionales Handeln beruht auf praktischen Schlüssen – mit dem Argument auf Basis der backward connection übereinstimmt. Daher bespreche ich beide gemeinsam. Vgl. Walker 1989, 673, FN. 12.

Dies wird aber als unmöglich abgelehnt, oder der Grad der dabei involvierten Irrationalität als zu hoch angesehen, um übliche Fälle von Willensschwäche erläutern zu können. Theorien, die davon ausgehen, dass die Person tatsächlich zwei einander widersprechende Urteile fällt und aufrechterhält, bedienen sich daher einer Selbsttäuschungstheorie, um diese Möglichkeit einsichtig zu machen.

David Pears sieht diesen rückwärts gerichteten Zusammenhang zwischen intentionalen Handlungen und besseren Urteilen bei Davidson am Werk, als Umkehrung seiner beiden Prinzipien:<sup>360</sup> Wenn eine Person etwas intentional tut, dann strebt sie diese Handlung mehr an als alternative Handlungen, und wenn sie eine Handlung mehr anstrebt als andere dann beurteilt sie diese auch als besser gegenüber anderen. Als Gründe gegen die Gültigkeit der backward connection führt Pears folgende an:

a) eine Person könne aus Gewohnheit handeln und in diesen Fällen sei es unplausibel, der Person zu unterstellen, sie habe eine besonders starke Strebung nach dieser Handlung.<sup>361</sup> Dagegen wende ich ein, dass die Strebung für eine Handlung nicht besonders stark, sondern lediglich stärker als alternative Strebungen sein müsste, um dem Prinzip von Davidson zu genügen.

b) eine Person könne etwas intentional tun, auch ohne der Handlung einen Wert beizumessen (idle curiosity).<sup>362</sup> Diese Fälle können aber auch so interpretiert werden, dass die Person die Handlung als zumindest interessanter als alternative Handlungen einschätzt, und mithin ein entsprechendes Werturteil vorliegt. Diese beiden Gegenbeispiele sind meiner Meinung nach nicht besonders plausibel, da sie auch leicht anders interpretiert werden können.

c) man müsse zwischen starken und schwachen Werturteilen unterscheiden: die backward connection gelte nur für schwache Werturteile, nicht jedoch für starke, mit denen Davidson operiere.<sup>363</sup> Schwache Werturteile können Akteuren auf Basis ihres Verhaltens zugeschrieben werden, starke Werturteile sind im Einklang mit den langfristigen Interessen des Akteurs, entsprechen demnach dem rationalen Standpunkt des Akteurs.

„A value judgement in this strong sense expresses a special kind of preference, based on one's own long-term interests or perhaps on other people's interests, and contrasted both with a desire based on a calculation of immediate pleasure and with a sudden yen for a cup of cocoa. This is the voice of reason and

---

<sup>360</sup> Vgl. Pears 1982, 40ff.; Pears 1984, 183–219; Kap. II.2.2.

<sup>361</sup> „A human agent is not an inert system always needing the push of wanting most. The inertia often works in the opposite way: the system continues of its own momentum until it runs into the opposition of wanting less than most.” (Pears 1982, 41)

<sup>362</sup> „An agent can do something because he wants to do it most without setting any value either on the particular thing that he does or on the general policy of doing what he most wants to do. A clear example of this kind is indulging idle curiosity.” (Pears 1982, 41)

<sup>363</sup> „It will be conceded that, with a small qualification, the backward connection is necessary when the weak concept of a 'value-judgement' is used, but denied that it is necessary when the strong concept is used.” (Pears 1984, 196f.)

the argument will be that, in many cases of *akrasia*, the desire that rebels against reason's edict does not even pretend to speak with the voice of reason." (Pears 1984, 196)

Nach dieser Unterscheidung gibt es demnach einen trivialen Sinn von ‚Werturteil‘ im Sinne einer enthüllten Präferenz, die dem Akteur zugeschrieben werden kann alleine auf der Basis, dass die Person diese Handlung vollzogen hat. Dieses Modell korrespondiert mit basalen entscheidungstheoretischen Annahmen. Der erwartete Wert der Handlung aus Sicht der Person muss diesem Verständnis nach größer gewesen sein als der erwartete Wert alternativer Handlungen, und er ergibt sich aus der Multiplikation des Werts des Handlungsergebnisses mit der Wahrscheinlichkeit des Eintreffens des Ergebnisses aufgrund der Handlung aus der Perspektive der Person. Wenn die Präferenzen der fraglichen Person einigen grundlegenden Axiomen genügen, so lässt sich ihr Handeln als Maximierung des Erwartungswerts deuten.<sup>364</sup> Auch nach dieser Formel könnte der Umstand, weshalb eine Person eine bestimmte Handlung vollzogen hat, an mehreren Faktoren liegen: a) an der Präferenzstruktur der Person; b) an der Einschätzung der Wahrscheinlichkeiten durch die Person. Diese Faktoren lassen sich genau bestimmen, indem der Person Wetten angeboten werden mit bekanntem Wahrscheinlichkeitsgrad.<sup>365</sup> Anhand der so gestalteten Wetten, denen gegenüber eine Person indifferent ist, lassen sich die beiden Faktoren, die Wertungen der Person und ihre Einschätzung der Wahrscheinlichkeit, bestimmen. Diese Methode funktioniert nur dann, wenn die Präferenzstruktur und die Einschätzung der Wahrscheinlichkeiten einigen basalen Axiomen genügen.<sup>366</sup> Auch für Davidson lassen sich schwache Werturteile für jede Art von Pro-Einstellung konstruieren. Der bloße Wunsch nach etwas wäre demnach repräsentierbar durch ein entsprechendes Werturteil, wobei das Werturteil die konative Einstellung *ausdrückt*. Die frühe Theorie von Davidson operiert dabei mit einem eliminativen Begriff von ‚Intention‘: eine Intention ist dort nur ein synkategorematischer Ausdruck ohne referenzielle Funktion.<sup>367</sup> Die spätere Theorie operiert dagegen mit einem reduktionistischen Begriff von ‚Intention‘: Intentionen sind spezielle Arten von nicht-konditionalen Werturteilen. Diese Theorie ist reduktionistisch, da Intentionen auch hier lediglich eine Spezies von Werturteilen sind, und keine eigenständige Art psychischer Einheiten bildet. Eine nicht-reduktionistische Theorie von ‚Intentionen‘ wäre eine, die Intentionen als eigenständige, nicht-reduzierbare psychische Entitäten auffasste.

---

<sup>364</sup> Vgl. Velleman 1993; Resnik 1987, ch. 4.

<sup>365</sup> Diese Methode geht zurück auf Ramsey, Morgenstern und von Neumann. Vgl. Ramsey 1926; von Neumann & Morgenstern 1944. Für Details s.u.

<sup>366</sup> So muss z.B. die Summe aller Wahrscheinlichkeiten genau 1 ergeben, und die Präferenzstruktur muss transitiv sein: Wenn A wertvoller ist als B, und B wertvoller als C, dann ist auch A wertvoller als C. Erfüllt die Präferenzstruktur diese Axiome, so lässt sich zeigen, dass eine hinsichtlich linearer Transformationen eindeutige Nutzenfunktion existiert, welche den Erwartungsnutzen der Optionen entspricht. Vgl. Velleman 1993.

<sup>367</sup> Vgl. Kap. II.2

Pears argumentiert für die Notwendigkeit der Gültigkeit der backward connection für *schwache* Werturteile folgendermaßen:

„Now when an agent does something intentionally, two conditions have to be met: he must know what he is engaged in doing, even if it turns out to be a failure, and he must be engaged in doing it because he wants to do it. This yields the quick result that every intentional action must be accompanied by a supportive value-judgement. For when the value judgement is a weak one, it merely has to express the desire from which the action issued.” (Pears 1984, 197)

Die erste Bedingung, die Wissensbedingung, werde ich hier nicht weiter diskutieren. In der einen oder anderen Form wird der Person zumindest ein Bewusstsein, wenn schon kein Wissen davon zugeschrieben, was sie gerade tut. Die zweite Bedingung formuliert einen Zusammenhang zwischen ‚intentional handeln‘, ‚etwas anstreben‘ und der Möglichkeit, ein entsprechendes Werturteil zumindest konstruieren zu können. Bei jeder Handlung, so diese These, lässt sich trivialerweise urteilen, dass die Präferenz des Akteurs für diese Handlung höher gewesen ist als für alle anderen. Dies lässt sich auch so verstehen, dass die Person das tut, wozu sie am meisten motiviert ist. Diese Motivation lässt sich durch ein Werturteil ausdrücken. Die Frage, ob motivationale Willensschwäche möglich ist, betrifft dann die Frage, ob jemand explizit ein Werturteil fällen kann, das gegen die willensschwache Handlung spricht, und dennoch so handeln, dass die Motivation für sein Handeln durch ein Werturteil ausgedrückt werden kann, das für die willensschwache Handlung spricht und damit dem ursprünglichen besten Werturteil widerspricht.

Zur Entscheidung dieser Frage muss der Standpunkt desjenigen, der entsprechende Werturteile konstruiert, berücksichtigt werden. Dieser könnte die Präferenzstruktur der Person über das Anbieten geeigneter Trade-Offs feststellen und gleichzeitig damit die Einschätzungen bestimmter Wahrscheinlichkeiten durch den Akteur bestimmen.<sup>368</sup> Nun verhält sich die Person aber willensschwach, und wird daher folgendes Muster zeigen: Obwohl sie bei vielen Situationen eine gewisse Präferenzstruktur zeigt, und auch über geeignete Einschätzungen der Wahrscheinlichkeiten für das Eintreten entsprechender Ereignisse verfügt, verhält sie sich nicht gemäß der durch die Theorie vorausgesagten Weise: sie tut nicht das, was ihrem besten Urteil – das durch den Interpreten konstruierte – entspricht, sondern etwas anderes. Zwei Möglichkeiten bieten sich nun dem Interpreten dieses Verhaltens an (der zugleich der Akteur selbst sein könnte): 1) die Präferenzstruktur und/oder die Wahrscheinlichkeitseinschätzungen waren doch anders, als sie vermeintlich festgestellt wurden; 2) die Person verfügt über eine Präferenzstruktur und/oder über

---

<sup>368</sup> Vgl. die Methode von Ramsey in Ramsey 1926. Vgl. auch Resnik 1987. Angenommen, eine Person präferiert A gegenüber B, und B gegenüber C. Dann lässt sich die Nutzenfunktion gegenüber diesen drei Objekten skalieren, indem der Person eine Lotterie angeboten wird, bei der sie zu einer bestimmten Wahrscheinlichkeit entweder A gewinnt oder C. Die Wahrscheinlichkeit, bei der die Person dieser Lotterie gegenüber indifferent ist in Vergleich zu Objekt B ergibt den relativen Abstand des Nutzens von B gegenüber A respektive C für die Person.

Wahrscheinlichkeitseinschätzungen, die nicht den Axiomen der Theorie entsprechen, demnach für diese Theorie nicht ‚rational‘ sind und daher auch nicht durch diese Theorie vorhergesagt werden können. Theorien, welche Willensschwäche in einer starken Form nicht zulassen, empfehlen dem Interpreten, stets Option 1 zu wählen. Der Interpret hat sich geirrt, die Theorie ist anwendbar, nur die Daten waren falsch. Theorien, welche Option 2 wählen, lassen die Theorie auch, wie sie ist, schränken ihre Anwendbarkeit jedoch auf rationale Akteure ein, und versuchen kausale Theorien anzubieten, welche das Abweichen von diesen rationalen Strukturen zumindest teilweise erklärbar machen. Welche dieser beiden Strategien besser ist, ist eine schwierige Frage. Zu bemerken ist folgendes: Während Anhänger der Option 1 sich auf ein behavioristisches Kriterium für die Zuschreibung von mentalen Einstellungen (Präferenzstrukturen und Wahrscheinlichkeitseinschätzungen) verlassen, sind Anhänger von Option 2 nicht auf ein solches Kriterium festgelegt – das bessere Urteil ist eines, das der Person normalerweise direkt epistemisch zugänglich ist und nicht erst induktiv erschlossen werden muss. Damit können Anhänger der zweiten Option zumindest einen Teil des Verhaltens von willensschwachen Akteuren direkter modellieren, nämlich über die Selbsteinschätzung der Person und ihrer Berichte über ihre eigene Präferenzstruktur. Für Anhänger der Theorie 1 sind solche Berichte allesamt falsch und Ausdruck von Selbsttäuschungsprozessen, falls sie sich nicht mit der in Handlungen geäußerten Präferenzstruktur decken. Willensschwäche wäre somit auch angesichts der Gültigkeit der backward connection auf zwei Weisen möglich: die willensschwache Person handelt nach einem schwachen Werturteil, das im Widerspruch mit einem von ihr gehegten starken Werturteil steht. Oder die willensschwache Handlung fällt nicht in den Anwendungsbereich rationaler Entscheidungstheorie aufgrund der irrationalen Struktur der zugrundeliegenden Präferenzen des Akteurs.

#### **IV.2.2 Forward connection**

Die forward connection formuliert die entgegengesetzte Richtung der backward connection, vom Werturteil zur Handlung.<sup>369</sup> Es wird damit behauptet, dass (bestimmte) Werturteile hinreichend sind für intentionale Handlungen. Das entsprechende Konditional lautet: Wenn eine Person eine Handlung als beste bzw. als besser gegenüber allen anderen beurteilt, dann handelt sie entsprechend intentional (und ungezwungen).<sup>370</sup> Das Argument gegen Willensschwäche auf Basis

---

<sup>369</sup> Dieses Prinzip hat bereits in Kap. III.5 eine zentrale Rolle gespielt. Hier wird es unter dem Gesichtspunkt der Intentionalität der Handlung diskutiert.

<sup>370</sup> Bei Davidson wird dies über die beiden Prinzipien P1 und P2 ausgedrückt, welche durch den Begriff des ‚anstrebens‘ (‚want to do more‘) zusammenhängen. Vgl. Kap. II.2.2. Problematisch an dieser Interpretation von Davidson ist, dass hier gar kein hinreichendes Kriterium für das Stattfinden von intentionalen Handlungen formuliert wird, da in P1 als zusätzliche Bedingung angegeben wird, dass die Person eine der beiden Handlungen intentional tut. Die forward connection wird hier demnach nur eingeschränkt behauptet. Doch reicht die Einschränkung hier für das Problem der Willensschwäche aus, da diese darin besteht, dass eine Person intentional

der forward connection, komplementär zum vorigen Argument von Walker, kann so formuliert werden:

- (P1) Wenn S eine Handlung F als beste beurteilt, dann tut S intentional F. (forward connection)
- (P2) Wenn S willensschwach handelt, dann beurteilt S Handlung F als beste und tut dennoch non-F<sup>371</sup> intentional. (Definition von motivationaler Willensschwäche)
- (P3) Es ist (fast) unmöglich, dass S zugleich F und non-F intentional tut.
- (K) Also: Willensschwaches Handeln ist unmöglich.

Das Argument basiert darauf, dass Prämisse 3 meist von allen unwidersprochen vorausgesetzt wird, also dass niemand zwei einander ausschließende Handlungen gleichzeitig intentional tun kann. Die Absurdität dieser reductio liegt demnach im Handeln selbst, im Tun zweier einander ausschließender Handlungen, nicht in der Annahme, zugleich einander widersprechende Urteile zu fällen wie bei dem Argument auf Basis der backward connection. Ich denke, dass diese argumentative Differenz zwischen der Plausibilität der Prämissen 3 in den beiden Argumenten (backward und forward) dazu führt, dass zumeist auf Basis der forward connection gegen Willensschwäche argumentiert wird, und nicht auf Basis der backward connection.<sup>372</sup> Kurz gesprochen: es ist absurder anzunehmen, jemand tue zugleich und intentional zwei einander ausschließende Handlungen, als jemand glaube zwei einander widersprechende Urteile. Unabhängig davon, ob es überhaupt möglich ist, zugleich zwei verschiedene Handlungen zu vollziehen, wäre diese ‚Lösung‘ für das Phänomen der Willensschwäche unbefriedigend. Dies deshalb, da davon ausgegangen werden sollte, dass die willensschwache Person vor einer echten Wahl von einander ausschließenden Alternativen steht, bevor sie handelt. Dasselbe anhand eines Beispiels: Wäre es für Markus möglich, zugleich an den Strand zu gehen und seine Arbeit zu erledigen, dann käme gar kein für Willensschwäche charakteristischer Konflikt auf – er ginge einfach zum Strand *und* erledigte dort seine Arbeit. Diese Lösung käme demnach einer Auflösung des Phänomens gleich, und diente nicht dazu, das Phänomen der Willensschwäche zu erläutern.

---

gegen ihr bestes Urteil handelt. Da sie etwas Intentionales tut, trifft P1 zu, und damit greift die forward connection auch bei Davidson.

<sup>371</sup> Allgemein ist nicht ganz klar, was ‚non-F tun‘ eigentlich meint. Hier soll es lediglich bedeuten, dass S eine Handlung tut, die verschieden ist von F.

<sup>372</sup> Diese Asymmetrie wird, trotz der ansonsten detaillierten Argumentation, von Pears nicht explizit gewürdigt: „Philosophers who agree with Socrates’ rejection of conscious last-ditch *akrasia* have relied on the necessity of the forward connection between valuing and doing more often than they have relied on the necessity of the backward connection between doing and valuing. Their reason for this preference may have been the idea that valuation derives greater effectiveness from the forward connection than from the backward one. If that is so, it is puzzling, because it is hard to see how there could possibly be any difference between the effectiveness of the two connections.” (Pears 1984, 220)

Charakteristischerweise kranken auch die Lösungsversuche auf Basis der backward connection an dem Problem, den Willensschwäche inhärenten Konflikt aufzulösen anstatt ihn zu erläutern: Wenn die Person genau das tut, was sie in der Situation für das Beste hält – da sie nicht zwei einander widerstreitende Aussagen zugleich für wahr halten kann, ändert sie ihr bestes Urteil vor der entsprechenden Handlung – ist unklar, worin genau das Problem der Willensschwäche bestehen sollte. Der Konflikt wird entsprechend als diachroner Konflikt erläutert, als Urteilswechsel oder Meinungsumschwung. Aber auch im Falle der forward connection könnte der Konflikt diachron interpretiert werden, z.B. als Aufschub des Handelns bzw. Prokrastination. Die Person vollzieht nicht gleichzeitig einander ausschließende Handlungen, sondern unterbricht die eine Handlung zugunsten einer anderen, ohne die andere abzuschließen, oder sie schiebt den Handlungsbeginn immer weiter in die Ferne, ohne je anzufangen. Letzteres spricht übrigens dagegen, das forward connection Argument als stärker als das backward connection Argument anzusehen.<sup>373</sup>

Folgende Positionen unterschreiben unter anderem die Gültigkeit der forward connection:

a) Positionen, die evaluative Urteile mit motivationalen Zuständen eng verknüpfen: Einige Varianten dieser Positionen habe ich bereits in Kap. III diskutiert.

b) Positionen, die Intentionen als bestimmte Art von Werturteil explizieren: Diese Positionen sind Unterarten von a), die den motivationalen Zustand als Intention spezifizieren. Von Intentionen wird weiterhin behauptet, dass für diese die forward-connection zutreffe: Intentionen seien so definiert, dass eine Person, wenn sie eine Intention hat, etwas zu tun, sie es unter normalen Umständen auch tut. Die für Willensschwäche genannten Umstände seien außerdem normal, daher könne niemand intentional willensschwach handeln. Die Frage bleibt, weshalb Intentionen als bestimmte Arten von Werturteilen verstanden werden sollten.

c) Positionen, welche eine teleologische Handlungserklärung vertreten. Diese erklären Handlungen über den Hinweis auf das Ziel, das der Handelnde durch sein Handeln verwirklichen möchte. Dieses Ziel wird dabei als gut charakterisiert, zumindest aus Sicht des Handelnden. Die Idee ist demnach, dass das formale Ziel jeder Handlung die Realisierung eines Gutes ist, und die Angabe dieses Gutes die entsprechende Handlung rationalisiert und erklärt.<sup>374</sup>

Walker gibt verschiedene Möglichkeiten an, wie das forward connection Argument kritisiert werden kann.<sup>375</sup>

---

<sup>373</sup> Vgl. dazu Pears 1984, 221ff.

<sup>374</sup> Zur Kritik an diesem Prinzip, oft bezeichnet als ‚guise of the good‘ oder ‚sub specie boni‘, vgl. Velleman 1992; Stocker 1979.

<sup>375</sup> Vgl. Walker 1989, 667ff.

1) Die Beschreibung des Phänomens der Willensschwäche als Handeln wider das beste oder bessere eigene Urteil wird zurückgewiesen. Willensschwäche bestehe nicht in Handlungen wider das beste Urteil, sondern in Handlungen wider die eigene Vernunft. Diese Kritik sieht Handlungen wider das eigene beste Urteil nicht als hinreichendes und/oder notwendiges Kriterium für willensschwache Handlungen an.<sup>376</sup> Frank Jackson verfolgt diese Strategie in seiner Behandlung von Willensschwäche.<sup>377</sup> Das Handeln der willensschwachen Person lässt sich genauso wie andere Handlungen entscheidungstheoretisch darstellen: die Person setzt die Handlung, für die sie zum Zeitpunkt der Handlung die größte Präferenz hat. Mithin genügt sie der forward connection und der backward connection. Die Präferenz ergibt sich aus der Summe des Produkts der Wertung eines Sachverhalts mit der zugehörigen bedingten Wahrscheinlichkeit des Eintretens dieses Sachverhalts gegeben andere Sachverhalte. Der Unterschied zwischen willensschwachen und willensstarken Personen liegt in der Zuschreibung von Werturteilen *vor* der entsprechenden Handlung und der dynamischen Entwicklung dieser Urteile auf Basis neuer Informationen. Ein Beispiel: Ein überzeugter Raucher präferiert vor dem Eintreten des Drangs nach einer Zigarette, dass er raucht, sollte sich der Drang einstellen. Ein überzeugter Nichtraucher präferiert das Gegenteil. Beide handeln nicht willensschwach: Tritt der Drang nach einer Zigarette beim überzeugten Raucher auf, so raucht dieser, tritt er beim überzeugten Nichtraucher auf, so raucht dieser nicht. Der eine präferiert es, den Drang durch Rauchen zu stillen, der andere, diesen nicht zu stillen. Die Information, dass der Drang auftritt, ändert nicht die konditionale Bewertung der Präferenz gegenüber diesen beiden Sachverhalten. Die willensschwache Person ändert jedoch nach Jackson ihre Bewertung, und diese Änderung ist nicht alleine erklärbar durch Rückgriff auf geänderte Informationen, wobei Information hier beschränkt wird auf Aussagen über bedingte Wahrscheinlichkeiten. Die willensschwache Person präferiert, bevor der Drang eintritt, nicht zu rauchen, sollte er eintreten. Tritt der Drang jedoch auf, so präferiert sie, zu rauchen, um den Drang zu stillen, anstatt nicht zu rauchen und den ungestillte Drang auszuhalten. Entscheidend ist, dass die willensschwache Person dabei über keine neuen Informationen über den Zusammenhang der beiden Sachverhalte verfügt: Ihre Einschätzung der Wahrscheinlichkeit, dass Rauchen einen spezifischen Drang stillt, bleibt dieselbe und unterscheidet sie nicht von den anderen Typen (Raucher bzw. Nichtraucher).

„We can express the difference between the strong-willed smoker and the weak-willed one in terms of the distinction between conditional wants (discussed earlier) and conditionals about wants. Neither smoker has the conditional want to smoke if he has no craving; both are such that if they have the craving, then they

---

<sup>376</sup> Diese Lösung erinnert an die entsprechende Lösung für die backward connection zwischen schwachen und starken Werturteilen zu unterscheiden.

<sup>377</sup> „There is a long tradition according to which weak-willed action is action *not* in accord with the agent's reason – ‘passion pushes reason from its throne’; thus my suggestion is that weak-willed action is action arising from wants and desires that have *not* evolved according to the dictates of the agent's reason, as just defined.” (Jackson 1984, 14)

want to smoke. But they differ in that when they have no craving, the strong-willed smoker nevertheless has the conditional want to smoke if he craves to (because his wants evolve by conditionalization), whereas the weak-willed smoker does not. He wants *not* to smoke if he craves to – although when the time comes, he will want to smoke.” (Jackson 1984, 16)

Ich gebe folgend eine Tabelle, um die Unterschiede der Bewertungsfunktion der drei Typen nach Jackson zu illustrieren.<sup>378</sup> ‚Vorher‘ und ‚nachher‘ bezeichnet dabei den Zeitpunkt des Auftretens des Drangs nach einer Zigarette. V (Rauchen & Drang) bezeichnet z.B. die Bewertung für den Sachverhalt, dass geraucht wird und dennoch ein ungestillter Drang besteht. Während der willensschwache Raucher zuerst dieselbe Bewertungsfunktion wie der Nichtraucher aufweist, ändert sich diese unter Einfluss des Auftretens des Drangs. Dieses Beispiel arbeitet unter der idealisierenden Voraussetzung, dass die Akteure den Wert des Rauchens einzig und allein im Hinblick auf seine Fähigkeit attraktiv finden, einen entsprechenden Drang zu stillen. Ohne diese Voraussetzung könnte der Raucher auch rauchen, ohne gegenwärtig einen Drang danach zu verspüren – er könnte einfach aus Freude rauchen, oder aus sozialen Motiven o.ä. Natürlich unterscheiden sich Nichtraucher und Raucher dahingehend, dass es sehr unwahrscheinlich ist, dass der Nichtraucher einen Drang verspürt, der durch Rauchen gestillt werden könnte, aber dies sollte hier nicht weiter stören.

	<b>Willensstarker und überzeugter Raucher</b>	<b>Willensstarker und überzeugter Nichtraucher</b>	<b>Willensschwacher Raucher und überzeugter Nichtraucher</b>	
<b>Bewertungsfunktion</b>	<b>vorher und nachher</b>	<b>vorher und nachher</b>	<b>vorher</b>	<b>nachher</b>
<b>V (Rauchen &amp; Drang)</b>	niedrig	niedrig	niedrig	niedrig
<b>V (Rauchen &amp; kein Drang)</b>	hoch	niedrig	<b>niedrig</b>	<b>hoch</b>
<b>V (kein Rauchen &amp; Drang)</b>	niedrig	hoch	<b>hoch</b>	<b>niedrig</b>
<b>V (kein Rauchen &amp; kein Drang)</b>	hoch	hoch	hoch	hoch

Die Einschätzung der Wahrscheinlichkeiten ist nach Jackson dieselbe für alle drei Typen. P (Drang/Rauchen) heißt hier die bedingte Wahrscheinlichkeit, dass ein ungestillter Drang besteht, obwohl geraucht wird. Die Tabelle gibt wieder, dass es die Akteure für wahrscheinlich halten, dass Rauchen einen spezifischen Drang zuverlässig befriedigt (hier idealisierend mit Wahrscheinlichkeit 1 bewertet), und dass nur Rauchen diesen erfüllt (Wahrscheinlichkeit bei idealisierend 0). Bevor der Drang auftritt, halten es die Akteure für unwahrscheinlich, dass wenn sie nicht rauchen, sie einen entsprechenden Drang haben. Dies ändert sich, wenn der entsprechende Drang auftritt: dann ist es wahrscheinlich, dass der Drang bestehen bleibt, wenn nicht geraucht wird, und unwahrscheinlich, dass er einfach verschwindet, wenn nicht geraucht

<sup>378</sup> Vgl. Jackson 1984, 16.

wird. Das Auftreten des Drangs wird von Jackson als Information für das Vorliegen eines bestimmten Sachverhaltes modelliert.

Wahrscheinlichkeitsfunktion	vorher	nachher
P (Drang/Rauchen)	0	0
P (kein Drang/Rauchen)	1	1
P (Drang/nicht Rauchen)	0	1
P (kein Drang/nicht Rauchen)	1	0

Entscheidungstheoretische Aussagen hängen immer auch daran, wie die jeweilige Situation genau beschrieben wird, d.h. welche Entscheidungen aus Sicht der Person anstehen. Hier ist Jackson inkonsequent in seiner Behandlung der Wahrscheinlichkeiten. Denn einmal bezeichnet P (kein Drang/Rauchen) einfach den Umstand, dass wenn kein Drang vorliegt, auch die Wahrscheinlichkeit null ist, dass sie unter bestimmten Umständen wie Rauchen vorliegt. Tritt der Drang aber auf, so bedeutet dieselbe Aussage, dass der Drang verschwindet, sollte geraucht werden, und ist demnach eine Aussage über die Zukunft, nicht über die Gegenwart. Denn bei gegebenem Drang wäre die Wahrscheinlichkeit, dass kein Drang besteht, falls geraucht wird, eben auch null und nicht eins. Auf Basis dieser geänderten Interpretation ist es durchaus auch plausibel, der Person *vor* dem Eintreten des Dranges eine positive Wahrscheinlichkeit für den Zusammenhang zuzuschreiben, dass das Rauchen den Drang nach Zigaretten begünstigt, also dass P (Drang/Rauchen) größer ist als 0. ‚Drang‘ sollte daher nicht den Sachverhalt bezeichnen, gegenwärtig einen Drang zu haben, sondern in naher Zukunft einen solchen zu haben. Folgende Tabelle der Wahrscheinlichkeiten wäre dann durchaus plausibel:

Wahrscheinlichkeitsfunktion	vorher	nachher
P (Drang/Rauchen)	0,7	0,3
P (kein Drang/Rauchen)	0,3	0,7
P (Drang/nicht Rauchen)	0,3	0,7
P (kein Drang/nicht Rauchen)	0,7	0,3

Die Annahme, dass die Einschätzung der Wahrscheinlichkeit für alle drei Typen dieselbe ist, ist problematisch. So könnten Nichtraucher generell davon ausgehen, dass der Drang nach einer Zigarette auch ohne zu rauchen einfach aufhören wird – also P (kein Drang/nicht Rauchen) höher ist als null –, selbst wenn der Drang vorliegen sollte. Der Unterschied beträfe dann nicht nur die Präferenzfunktion, sondern auch die Wahrscheinlichkeitsfunktion: beide glauben daran,

dass Rauchen ein effektives Mittel ist, um den Drang nach Zigaretten zu stillen – also  $P$  (kein Drang/Rauchen) ist nahe eins –, aber Raucher glauben außerdem an die Ausschließlichkeit dieses Mittels – also  $P$  (kein Drang/nicht Rauchen) ist nahe null.

Die gesamte Präferenzfunktion ergibt sich aus der Summe der Produkte der individuellen Bewertungen der Sachverhalte mit deren bedingten Wahrscheinlichkeiten:

$$V(\text{nicht Rauchen}) = V(\text{nicht Rauchen \& Drang}) \times P(\text{Drang/nicht Rauchen}) + V(\text{nicht Rauchen \& kein Drang}) \times P(\text{kein Drang/nicht Rauchen})$$

$$V(\text{Rauchen}) = V(\text{Rauchen \& Drang}) \times P(\text{Drang/Rauchen}) + V(\text{Rauchen \& kein Drang}) \times P(\text{kein Drang/Rauchen})$$

Die willensschwache Person ändert ihre Präferenz gegenüber dem Rauchen, und diese Änderung liegt nach Jackson nicht an unterschiedlichen Einschätzungen von Wahrscheinlichkeiten zwischen den Typen – diese sind für alle drei dieselben –, sondern an einer Änderung der Präferenz dafür, einen ungestillten Drang auszuhalten und nicht zu rauchen, gegenüber der Präferenz, den Drang durch Rauchen zu stillen. In der Diskussion um die backward connection hinsichtlich der Unterscheidung schwacher und starker Werturteile habe ich festgestellt, dass es entscheidungstheoretisch nicht eindeutig sei, ob die willensschwache Person ihre Präferenzen oder ihre Einschätzung der Wahrscheinlichkeiten (subjektive Wahrscheinlichkeit) geändert hat. Nach Jackson sollte die Änderung jedoch lediglich die Präferenz betreffen, nicht die subjektive Wahrscheinlichkeit. Ich halte dagegen daran fest, dass dies nicht eindeutig feststellbar ist, und dass aus Sicht der Entscheidungstheorie nichts dagegen spricht, Willensschwäche als irrationale Änderung der Wahrscheinlichkeitseinschätzung der Person zu verstehen. Am Beispiel des Rauchens: hier wurde davon ausgegangen, dass sich die subjektive Wahrscheinlichkeit aller drei Typen nicht ändert. Es könnte jedoch auch sein, dass die willensschwache Person, bevor sie einen Drang nach einer Zigarette verspürt, mit hoher Wahrscheinlichkeit daran glaubt, dass dieser Drang auch ohne zu rauchen einfach verschwinden wird, wenn der Drang auftreten sollte – was auch tatsächlich der Fall ist, schließlich dauern die sogenannten Craving-Perioden von Nikotinsüchtigen normalerweise nicht sehr lange an. Wenn der Drang dann jedoch einsetzt, ändert sich diese Einschätzung, und die willensschwache Person glaubt, dass der Drang *nur* durch das Rauchen aufhören wird.<sup>379</sup>

Angenommen, die Präferenzstruktur des willensschwachen Rauchers sieht folgendermaßen aus, wobei das ‚>‘ als ‚ist präferiert gegenüber‘ gelesen werden kann:

---

<sup>379</sup> Dies unterscheidet den Fall übrigens von dem Durst-Beispiel, das Jackson zuvor bespricht (vgl. Jackson 1984, 14f.). Bei Durst ist es tatsächlich plausibel anzunehmen, dass vor und nach Einsetzen des Durstes die Einschätzung der Wahrscheinlichkeit, durch Trinken von Flüssigkeiten und *nur* durch das Trinken von Flüssigkeiten den Durst zu stillen, bei allen drei Typen gleich bleibt – abgesehen davon, dass Willensschwäche hier wohl ein Sonderfall wäre, z.B. im Falle eines Durststreiks o.ä.

- V (nicht Rauchen & kein Drang)
- > V (Rauchen & kein Drang)
- > V (nicht Rauchen & Drang)
- > V (Rauchen & Drang)

Zusätzlich nehme ich an, diese Präferenzstruktur wird durch die Werte 20, 15, 10 und 5 repräsentiert, ohne diesen Zahlen eine besondere Bedeutung zu verleihen. Diese Präferenzstruktur ist hier dieselbe wie die eines Rauchers, d.h. die willensschwache Person zieht es vor, zu rauchen statt einen ungestillten Drang auszuhalten, aber dennoch denkt sie, dass nicht zu rauchen und keinen Drang zu haben besser ist als die übrigen Alternativen. Nun ergeben sich folgende Möglichkeiten, wenn der Drang einsetzt:

a) die Person glaubt, dass nur das Rauchen ihren Drang stillt, P (kein Drang/nicht Rauchen) ist demnach gleich null. Die Person raucht, da V (nicht Rauchen) kleiner ist als V (Rauchen). In der Formel:

$$V(\text{nicht Rauchen}) = 10 \cdot 1 + 20 \cdot 0 = 10$$

$$V(\text{Rauchen}) = 5 \cdot 0 + 15 \cdot 1 = 15$$

b) die Person glaubt, dass, auch wenn sie nicht raucht, der Drang vergeht, d.h. dass P (kein Drang/nicht Rauchen) größer ist als null. Wenn dieser Glaube entsprechend hoch ist – in diesem Beispiel jeder Wert höher als 0,5 –, so wird die Person nicht rauchen. In der Formel, wobei für die Wahrscheinlichkeit 0,6 eingesetzt wird:

$$V(\text{nicht Rauchen}) = 10 \cdot 0,4 + 20 \cdot 0,6 = 16$$

$$V(\text{Rauchen}) = 5 \cdot 0 + 15 \cdot 1 = 15$$

Nun könnte die Person, bevor sie den Drang nach einer Zigarette verspürt, das hypothetische Urteil fällen, dass sie, falls solch ein Drang auftritt, die Wahrscheinlichkeiten im Sinne von b statt von a einschätzen sollte. Diese Überzeugung ändert sie jedoch ohne plausiblen Grund, sobald sie den Drang nach einer Zigarette verspürt. Die Theorie müsste demnach nicht nur irrationalen Präferenzwandel, sondern auch irrationalen Überzeugungswandel erfassen. Problematisch ist, dass dadurch die klare Eingrenzung von Willensschwäche auf Formen des Präferenzwandels innerhalb der Theorie verlorenght.

Ein viel grundlegendes Problem an der Theorie von Jackson ist, dass *jede* Änderung der Bewertung, die nicht auf Änderungen der Einschätzung der bedingten Wahrscheinlichkeiten beruht, als irrational und willensschwach angesehen wird. Dagegen hat Christopher Corder eingewandt, dass es Fälle von rationalen Wertewandel geben könne, der nicht willensschwach sei,

und daher durch diese Theorie nicht erfasst werde.<sup>380</sup> Als Beispiel führt er die moralische Konversion einer Person an, die einen radikalen Wertewandel ihrer religiösen Einstellungen vollzieht, ohne dabei auf neue Informationen zurückzugreifen.

„Action issuing from this change of value ought, on Jackson's account, to be akratic. But clearly it is not. The proper way to describe this person, in the above situation, is not as weak-willed, but as having undergone a radical change of values (whether or not one we approve of is another matter). The person himself, after all, will not speak of his actions as weak-willed, nor of his change of values as contrary to reason. He will rather speak of the corruptness of his former values and the rightness of his present course. And on what basis, other than the defence of a philosophical theory, could we object to his speaking in this way? He feels no conflict within himself, nor any regret over his changed course; and it is surely grossly implausible to suppose that this convert must be self-deceived in his rejection of the idea that his actions as a convert are akratic.” (Cordner 1985, 276)

Doch es braucht keinen Verweis auf Fälle von *radikalem* Wertewandel, um ein Gegenbeispiel zur Theorie von Jackson zu konstruieren, denn *jede* Form des Wandels ursprünglicher Präferenzen ist nach dieser Theorie willensschwach. Einen möglichen Einwand gegen dieses negative Resultat formuliert Cordner: die Theorie beschreibe nur Willensschwäche als Resultat des Einflusses von körperlichen Begierden, normale Formen von Wertewandel seien daher nicht erfasst und demnach nicht willensschwach. Dagegen lassen sich aber Fälle konstruieren, in denen Wertewandel zumindest auf körperlichen Begierden beruhen, diese jedoch nicht als willensschwach beschrieben werden sollten.<sup>381</sup>

Ein Merkmal der Theorie von Jackson ist sicherlich, dass sie – wie andere bereits besprochene Theorien auch – Willensschwäche nur als diachrones Phänomen erfassen kann: zum Zeitpunkt der willensschwachen Handlung macht es keinen Sinn, dem Akteur ein besseres Urteil zuzuschreiben, dass gegen diese Handlung spricht.<sup>382</sup> Jackson erwähnt, dass seine Theorie nur eine Theorie subjektiver Wertungen, und keiner objektiven sei.<sup>383</sup> Möglicherweise könnte die Theorie erweitert werden, um auch Glaubenszustände der Person über objektive Wertungen zu repräsentieren. Dann würde es Sinn machen, der Person zugleich ein Urteil zuzuschreiben, dass ein Sachverhalt einem anderen objektiv vorzuziehen ist, ohne dass sich diese Bewertung mit der subjektiven Präferenzfunktion der Person gegenüber diesem Sachverhalt widerspiegelt. Dies geht jedoch über die engen Grenzen der Entscheidungstheorie hinaus.

2) Das Prinzip einer direkten Korrelation zwischen evaluativen Urteilen und der Stärke von Wünschen wird zurückgewiesen. Dies geschieht z.B. durch eine Unterscheidung verschiedener

---

<sup>380</sup> Vgl. Cordner 1985.

<sup>381</sup> Vgl. Cordner 1985, 276ff.

<sup>382</sup> „Jackson has temporalized a conflict which we naturally think of as engaged at the time of the akratic choice, though of course it may have stirred before then.” (Cordner 1985, 279)

<sup>383</sup> Vgl. Jackson 1984, 11. In anderen Worten: die Theorie handelt nur von motivationalen Gründen in einem internalistischen Sinn und nicht von normativen Gründen.

Arten von Wünschen.<sup>384</sup> So unterscheidet Elizabeth Rapaport zwischen bloßen Wünschen (*mere desires*) und rationalen, an Moral und Klugheit orientierten Wünschen (*rational desires*). Für letztere sei das Prinzip der *forward connection* gültig, es könne jedoch nicht ausgeschlossen werden, dass ein Akteur neben seinen rationalen Wünschen noch andere, bloße Wünsche hege, die sein Handeln bestimmen. Dies ist ähnlich zu dem Vorschlag im Rahmen der *backward connection* zwischen schwachen und starken Werturteilen zu unterscheiden.

3) Das Prinzip einer direkten Korrelation zwischen alles in allem besten Urteilen und der Motivation entsprechend zu handeln wird qualifiziert. So argumentiert Charles Taylor, dass Davidsons P2 zu eng gefasst sei, und es daher erweitert werden müsste zu:

„P2”. If an agent judges that it would be better to do *x* than to do *y* he ranks the doing of *x* by him on this occasion higher than he ranks the doing of *y* by him on this occasion.” (Taylor 1980, 516)

Er schlägt vor, das Verhältnis zwischen Urteil und Handlung als einen Bündelbegriff aufzufassen: die Kriterien dafür, eine Handlung einer anderen vorzuziehen, sind weiter und umfassen daher mehr als bloß den tatsächlichen Vollzug der Handlung.<sup>385</sup> Dagegen hat G. F. Schueler eingewandt, dass dieses Konzept dazu führe, Personen inkohärente Einstellungen zuzuschreiben. Die willensschwache Person erfüllt einige hinreichende Bedingungen dafür, dass sie *x* höher einschätzt als *y*. Zugleich erfüllt sie aber hinreichende Bedingungen dafür, *y* höher als *x* einzuschätzen, schließlich vollzieht sie Handlung *y* und nicht *x*.<sup>386</sup> In einer Antwort reagiert Taylor auf diesen Einwand mit einer Präzisierung seiner Position:

„I grant, however, that I expressed myself badly in saying [...] that a sufficient condition for the truth of the proposition specifying the agent's ranking is the satisfaction of some disjunction of conjunctions of conditions. What I should have said was that no single condition or set of conditions is either logically necessary or logically sufficient, since any condition or set of conditions may be absent provided some other condition or set of conditions is present, and any condition or set of conditions may be overridden by countervailing conditions, though (a) it is logically necessary that some conditions are satisfied and (b) in the absence of countervailing conditions the satisfaction of some (unspecified) number of conditions is in practice sufficient.” (Taylor 1984, 586)

Dies hat jedoch die problematische Konsequenz, dass der willensschwachen Person keines der beiden Urteile zugesprochen werden kann.<sup>387</sup> Damit kann das bessere Urteil, gegen das die willensschwache Person handelt, in dieser Theorie nicht repräsentiert werden, und somit auch nicht willensschwaches Handeln, wodurch das selbstgesteckte Ziel der Analyse scheitert.

---

<sup>384</sup> So hat Gary Watson zwischen einem evaluativen und einem motivationalen Sinn von ‚mehr wünschen‘ unterschieden. Vgl. auch Richman 1980; Pears 1984.

<sup>385</sup> Vgl. Taylor 1980 und 1984 über den Begriff ‚to rank higher than‘ als Cluster-Konzept, das durch verschiedene Konditionale erfüllt werden kann.

<sup>386</sup> „It [Taylors Prinzip P2] allows the possibility that the same agent (at the same time, etc.) ranks *x* higher than *y* and ranks *y* higher than *x*.” (Schueler 1983, 583)

<sup>387</sup> „In some, perhaps most, cases of psychological conflict it will be true neither that the agent ranks his doing *x* above his doing *y* nor vice versa.“ (Taylor 1984, 586)

### IV.2.3 Gründe-Theorie intentionalen Handelns

Der Einwand auf Basis einer reason theory oder einer Gründe-Theorie intentionalen Handelns besagt, dass die willensschwache Person keinen Grund hatte, willensschwach zu handeln. Da jedoch alle intentionalen Handlungen aus Gründen vollzogen werden, hätte die willensschwache Person demnach nicht intentional handeln können.<sup>388</sup> Das Argument lässt sich folgendermaßen verstehen:

- (P1) Wenn eine Person intentional handelt, dann handelt sie aus einem Grund.
- (P2) Wenn eine Person willensschwach handelt, dann handelt sie aus keinem Grund.
- (K) Also: Wenn eine Person willensschwach handelt, dann handelt sie nicht intentional.

P1 formuliert den Zusammenhang zwischen Gründen und intentionalen Handlungen, P2 ist eine Beschreibung von willensschwachen Handlungen als grundlos. In dieser Form ist noch nicht ersichtlich, weshalb die willensschwache Person keinen Grund für ihre Handlung haben sollte. Dafür muss klar sein, was überhaupt als Grund für eine Handlung fungieren kann, mithin, was praktische Gründe sind. Intuitiv ergibt sich ein Dilemma, wenn die Relation der willensschwachen Person zu ihrem Grund erläutert wird:

- a) die Person handelt aus einem Grund, der aus ihrer Sicht die Handlung rechtfertigt: Weshalb kann sie dann gleichzeitig zu einem Urteil gelangen, dass ihre Handlung auf Basis ihrer anderen Gründe nicht gerechtfertigt ist?
- b) die Person handelt nicht aus einem Grund, sondern lediglich auf Basis von etwas, was Ursache, aber nicht Grund der Handlung ist. Wie kann das Handeln der Person als intentional charakterisiert werden auf Basis einer Gründe-Theorie intentionalen Handelns?

Das Argument der Gründe-Theorie kann demnach folgendermaßen verfeinert werden:

- (P1) Wenn eine Person intentional handelt, dann handelt sie aus einem Grund.
- (P2) Wenn eine Person aus einem Grund handelt, dann sieht sie ihr Handeln als durch diesen Grund gerechtfertigt an.
- (P3) Wenn eine Person willensschwach handelt, dann sieht sie ihr Handeln nicht als durch den Grund gerechtfertigt an, aus dem sie handelt.
- (K) Also: Wenn eine Person willensschwach handelt, dann handelt sie nicht intentional.

Eine alternative Variante wäre:

- (P1) Wenn eine Person intentional handelt, dann handelt sie aus einem Grund.
- (P2) Wenn eine Person aus einem Grund handelt, dann sieht sie ihr Handeln als durch diesen Grund gerechtfertigt an.

---

<sup>388</sup> Vgl. die Diskussion dieses Argumentes bei Charles 1982–83.

- (P3) Wenn eine Person willensschwach handelt, dann sieht sie ihr Handeln durch den Grund gerechtfertigt an, aus dem sie handelt.
- (P4) Wenn eine Person ihre Handlung als durch den Grund, aus dem sie handelt, als gerechtfertigt ansieht, dann beurteilt sie die Handlung als besser oder als beste gegenüber ihren Alternativen.
- (K) Also: Wenn eine Person willensschwach handelt, dann beurteilt sie die Handlung als besser oder als beste gegenüber ihren Alternativen

Die erste Version führt dazu, der willensschwachen Person überhaupt abzusprechen, intentional zu handeln, während die zweite im Konflikt steht zur Beschreibung von Willensschwäche als Handeln wider das bessere Urteil. Folgende Möglichkeiten gibt es unter anderem, auf dieses Argument zu reagieren:

a) Zurückweisung der Gründe-Theorie intentionalen Handelns: Intentionale Handlungen sind nicht explizierbar als Handlungen aus Gründen. Dies impliziert, dass es intentionale Handlungen geben kann, die nicht aus Gründen geschehen.<sup>389</sup> Alternativen wären: 1) Handlungstheorien, die auf den Begriff der Lenkung oder Handlungskontrolle aufbauen;<sup>390</sup> 2) Handlungstheorien, die Handlungen als Reaktionen auf bestehende Tatsachen modellieren;<sup>391</sup> 3) Handlungstheorien, die intentionale Handlungen als Tätigkeiten, die durch eine Person intendiert werden, verstehen, wobei die handlungsleitende Intention aber nicht unbedingt in einem Zusammenhang zu Gründen stehen muss. Problematisch an diesen Ansätzen ist, dass die Gründe-Theorie eine Explikation dessen anstrebt, was es heißt, auf Basis einer handlungsleitenden Intention etwas zu tun. Eine alternative Theorie müsste daher darlegen, was sie unter einer handlungsleitenden Intention versteht, da der Begriff der Intention für alle diese Handlungstheorien wichtig ist. Das Konzept eines Handlungsgrundes ist jedoch auch auf eine Weise fundamental für *jede* Explikation einer Handlung, wie folgendes Argument zeigen kann:

- (P1) Die handlungsleitende Intention einer Person ist auf mindestens ein Ziel gerichtet, das durch die Handlung erreicht werden soll. (Teleologische Struktur von Handlungen)
- (P2) Die Angabe eines Zieles kann als minimaler Grund dafür dienen, diese Handlung zu tun.<sup>392</sup>
- (K) Also: Die handlungsleitende Intention ist auf etwas gerichtet, das als minimaler Grund für die Handlung dienen kann.

---

<sup>389</sup> Vgl. Bittner 2011 für eine mögliche Variante einer solchen Theorie.

<sup>390</sup> Vgl. Überlegungen von Frankfurt 1978.

<sup>391</sup> Vgl. Bittner 2011.

<sup>392</sup> Das ist verschieden von der Aussage, die Intention für eine Handlung selbst wäre schon ein Grund, eine Handlung zu tun. Dies ist nämlich äußerst umstritten, wie die Diskussion um ‚bootstrapping‘ zeigt. Vgl. Bratman 1987, ch. 2; Broome 1999.

Da das Konzept eines Handlungsgrundes in diesem Sinne fundamental ist, und daher auch innerhalb solcher alternativen Handlungstheorien eine gewisse Rolle spielen muss, wäre es einfacher, wenn auch willensschwache Handlungen als begründete Handlungen aufgefasst werden könnten, anstatt als völlig grundlos. Einen Hinweis, wie dies möglich ist, liefert die alternative Version des obigen Argumentes. Prämisse 4\* besagt, dass die Person die Handlung auch als besser als alle Alternativen beurteilen muss, wenn sie diese als gerechtfertigt ansieht. Nach Davidson heißt dies aber, ein *prima facie* Urteil über den Zusammenhang eines Grundes mit einer Handlung mit einem *Alles-in-allem*-Urteil über alle vorliegenden Handlungsgründe zu verwechseln.

b) Die willensschwache Person hat einen Grund zu handeln, auch wenn sie diesen als schlechter ansieht als andere Gründe. Dafür braucht es eine Theorie, die zeigt, wie Handlungsgründe schlechter sein können, ohne zugleich ihren Status als Grund zu verlieren.<sup>393</sup> Elizabeth Rapaport unterscheidet dafür zwischen *adäquaten* und *inadäquaten* Gründen.<sup>394</sup>

„The reason why a weak agent fails to do what he ought, given his principles and available information, is that he succumbs to the temptation to do what he in fact does. This is an *inadequate* reason for acting. In order for a reason to be adequate for acting, the action must be for the best, all things considered. An adequate reason is a justifying reason. That something is morally or prudentially best justifies doing it. If he does not have an adequate reason, he nonetheless has a reason for acting, which although justificationaly inadequate – hence, the term “inadequate reason” – was motivationally sufficient to induce him to do the action. He wanted to do what he did more than he wanted to do what he thought best.” (Rapaport 1973, 177)

Ein adäquater Grund ist ein Grund, der eine Handlung aus Sicht der Person rechtfertigt gemessen an Prinzipien des klugen oder des moralischen Handelns, welche die Person akzeptiert.<sup>395</sup> Inadäquate Gründe sind solche, welche für eine Handlung aus Sicht der Person sprechen, jedoch in Konflikt stehen zu anderen, ebenfalls für sich hinreichenden Gründen, welche aus Sicht der Person besser sind und für eine alternative Handlung sprechen.

Auch Susan L. Hurley entwickelt in *Natural reasons* (1989) eine Theorie, welche erlauben soll, dass praktische Gründe miteinander in Konflikt stehen können, ohne ihre rechtfertigende Funktion gänzlich zu verlieren – sich gewissermaßen auszulöschen. Dafür unterscheidet sie zwischen *prima facie* und *pro tanto* Gründen.

---

<sup>393</sup> Auch die Theorie von David Wiggins kann so verstanden werden: die Inkommensurabilität von Werten stellt nach ihm sicher, dass die willensschwache Handlung zumindest teilweise begründet ist. Vgl. Kap. III.4.1.

<sup>394</sup> Vgl. Rapaport 1973; Walker 1989, 662ff.

<sup>395</sup> „In light of what has been said, we must distinguish between a reason for acting and an adequate reason for acting. Any desire together with a belief that the means of gratification lies to hand may be an agent's reason for acting. But an agent will have an adequate reason for acting, if prudential considerations alone are at stake, only if his reason for acting is that, given his scheme of preferences and information available, the course of action he chooses is the best thing to do in the circumstances. When moral considerations enter in, the agent has an adequate reason for acting only if his reason for acting is that his course of action is the best thing to do, given the moral principles he accepts and the information available.” (Rapaport 1973, 175)

„The term “*prima facie*” seems appropriate for reasons that are like the reasons given by relational probabilities in that they can’t come into ultimate conflict with one another or be the operative reasons in cases of *akrasia*. *Prima facie* reasons are like rules of thumb, that give us reasons provisionally but may turn out not to apply when we learn more about the situation at hand, in which case they have no residual reason-giving force.” (Hurley 1989, 133)

„Conflicting reasons for action such as justice and kindness aren’t *prima facie* in the sense that, like rules of thumb, they seem to give reasons to do acts that would be just or kind, but may turn out not to when we learn more about the situation. One could do the kind act because kindness requires it, even in the face of one’s better judgment, namely, that all things considered one ought not to do the kind act but rather the just act. In doing so, one would be acting irrationally, in the way that the *akrates* acts irrationally. But one would still be acting for a reason, namely, because of the kindness of the act. The term “*pro tanto*” seems appropriate to describe such reasons, which might conceivably come into ultimate conflict with one another and can thus be the operative reasons in cases of *akrasia*.” (Hurley 1989, 133f.)

Die Idee ist demnach, dass praktische Gründe, also Gründe, etwas zu tun, bestehen bleiben, auch wenn die Person zu dem Urteil gelangt, dass sie anderen Gründen unterlegen wären. Die Strategie dieser Ansätze ist klar: die Rationalität der einzelnen Handlung für sich alleine betrachtet soll gewahrt bleiben, damit die Gründe Theorie des intentionalen Handelns anwendbar bleibt, während die Handlung insgesamt betrachtet dennoch nicht gerechtfertigt ist aus Sicht der Person. Willensschwache Handlungen sind also lokal rational, global jedoch irrational.

### IV.3 Was sind praktische Gründe?

Die entsprechenden Theorien lassen sich zumindest anhand dreier Fragestellungen klassifizieren. 1) Internalistische und externalistische Konzeptionen normativer praktischer Gründe;<sup>396</sup> 2) Psychologistische und anti-psychologistische Konzeptionen; 3) Wunschbasierte und wertebasierte Konzeptionen. Dabei nimmt der Abstraktionsgrad dieser Perspektiven jeweils ab.<sup>397</sup>

#### IV.3.1 Motivationaler Internalismus und Externalismus von praktischen Gründen

Motivationaler Internalismus von Gründen ist die These, dass praktische Gründe, also Gründe, eine Handlung zu tun, einen internen Zusammenhang zur Motivation des Akteurs aufweisen, die Handlung zu tun. Diese Beziehung bestehe notwendig. Außerdem ist die Beziehung intern, da die motivationalen Ressourcen der Person, die Handlung zu tun, Fakten der Person betreffen, so wie sie derzeit ist.

Motivationaler Gründe-Internalismus: Notwendigerweise: Wenn eine Person einen Grund hat, F zu tun, dann hat sie auch ein Motiv, F zu tun.

Motivationaler Gründe-Externalismus verneint eine solche interne, notwendige Verbindung zwischen normativen praktischen Gründen für Handlungen und der Motivation, diese zu tun. Ein Formulierung wäre daher:

---

<sup>396</sup> Nicht zu verwechseln mit der Diskussion um motivationalen Urteils-Internalismus/Externalismus über das Verhältnis von Urteilen zu Motiven in Kap. III.5.

<sup>397</sup> Vgl. auch Ammereller 2005; Dancy 2000, ch. 1.

Motivationaler Gründe-Externalismus: Es ist möglich, dass jemand einen Grund hat, F zu tun, und kein Motiv, F zu tun.

Nach internalistischen Positionen hat die Person nur dann einen Grund, etwas zu tun, wenn sie auch ein entsprechendes Motiv aufweist, es zu tun. Die willensschwache Person hat demnach nur dann einen Grund für ihre Handlung, wenn sie auch motiviert ist, entsprechend zu handeln. Problematisch für internalistische Positionen ist die Frage, weshalb die Person mehr oder stärkere Gründe gehabt hat, nicht willensschwach zu handeln. Es ist unklar, worauf sich innerhalb dieser Positionen die Berufung auf starke und schwache Gründe bezieht, wenn nicht auf die Stärke und Schwäche der zugrundeliegenden Motive, möglicherweise noch in Gemeinschaft mit den sonstigen Überzeugungen der Person. Wenn die Person nun stärker motiviert war, die willensschwache Handlung auszuführen, dann hat sie gemäß diesem Modell das getan, wofür sie auch die besten Gründe hatte. Damit verschärft sich das normative Problem, weshalb eine Person, die das tut, was ihren besten Gründen entspricht, irrational oder falsch handle.

Aus externalistischer Sicht kann die Person auch Gründe dafür haben, etwas zu tun, obwohl sie kein entsprechendes Motiv hat, es zu tun. Problematisch ist hier, wie die Person als jemand charakterisiert werden kann, die aus Gründen handelt. Denn laut der Charakterisierung von Willensschwäche spricht aus Sicht der Person zwar etwas für die willensschwache Handlung, aber stärkere Gründe sprechen gegen diese, was sich in dem besseren Urteil ausdrückt. Daher ist unklar, was angesichts dieses Umstandes überhaupt noch für die willensschwache Handlung spricht. Es braucht demnach eine Theorie von pro tanto Gründen, von Gründen, die bestehen bleiben, auch wenn andere Gründe sie aufwiegen und übertreffen.<sup>398</sup> Problematisch an dieser sehr abstrakten Charakterisierung des Problems ist, dass ohne eine geeignete Interpretation dessen, was ein Motiv ist, sich die internalistische und externalistische Sicht nicht klar voneinander unterscheiden.

#### **IV.3.2 Psychologistische und anti-psychologistische Theorien praktischer Gründe**

Für psychologistische Theorien sind Handlungsgründe Fakten betreffend der Psyche einer Person, während für nicht-psychologistische Theorien praktische Gründe normalerweise Fakten der jeweiligen Situation sind, in der sich eine Person befindet. Die Einschränkung ‚normalerweise‘ bedeutet, dass manche Gründe tatsächlich Fakten über die Psyche einer Person sein können.

„Die These des Psychologismus besagt, dass motivierende Gründe für Handlungen *psychische Entitäten* sind. Z. B. ist der motivierende Grund von Anton, den Regenschirm mitzunehmen, *dass er glaubt, dass es regnet*. Diese These ist von Antipsychologen kritisiert worden: Eine überzeugende Theorie der

---

<sup>398</sup> Vgl. zuvor die Diskussion um das Konzept von pro tanto Gründen bei Hurley 1989.

Handlungsmotivation sei nur möglich, wenn motivierende Gründe nicht-psychische Entitäten sind. Antons motivierender Grund sei, *dass es regnet*, und nicht, *dass er glaubt, dass es regnet*.“ (Mantel 2009, 2)

Für psychologistische Theorien sind psychische Zustände von Personen, nämlich Überzeugungen, Wünsche, Intentionen oder Kombinationen von diesen die Gründe, aus denen eine Person handelt und die deshalb auch ihr Handeln erklären können. Für antipsychologistische Theorien sind Gründe normalerweise keine psychischen Zustände von Personen, sondern Fakten über die jeweiligen Weltzustände oder Fakten über normative Verhältnisse zwischen verschiedenen Werten oder Normen.

Oftmals wird in diesen Zusammenhang eine Unterscheidung zwischen *motivierenden* und *normativen* Gründen eingeführt, deren genaue Interpretation jedoch auch Teil der Diskussion darstellt. *Motivierende* Gründe sind die Gründe aus Sicht der handelnden Person, welche die Handlung leiten und nach denen die Person ihr Handeln ausrichtet. Normative Gründe sind Gründe, welche das Handeln tatsächlich rechtfertigen können. Folgende Positionen werden gegenüber dieser Unterscheidung vertreten: a) Motivierende Gründe *sind* einfach normative Gründe aus Sicht der Person. Es gibt demnach keine verschiedenen Arten von Gründen, sondern nur verschiedene Perspektiven, über normative Gründe zu sprechen, nämlich einmal in ihrer motivationalen und handlungserklärenden Rolle (motivierende Gründe), ein anderes Mal in ihrer rechtfertigenden Rolle (normative Gründe).<sup>399</sup> Problematisch ist für diese Positionen die Erklärung von Handlungen, die auf falschen Überzeugungen basieren.<sup>400</sup> b) Normative Gründe und motivierende Gründe bezeichnen verschiedene Entitäten und sind daher strikt voneinander getrennt: motivierende Gründe sind psychische Tatsachen, normative Gründe sind normative Tatsachen.<sup>401</sup> Problematisch ist an diesen Positionen hauptsächlich, den Zusammenhang zwischen motivierenden und normativen Gründen zu erläutern, vor allem hinsichtlich der Frage, wie jemand aus guten, also normativen Gründen, handeln kann. c) Alle normativen Gründe sind motivierende Gründe, welche wiederum aus psychischen Tatsachen konstituiert sind – Wünsche, Überzeugungen oder Kombinationen aus diesen. Problematisch für diese Positionen ist, den dem Modell inhärenten rein instrumentellen Charakter von Rationalität zu verteidigen und dabei die Intuition der Universalität von Geltungsansprüchen von Gründen einzulösen oder plausibel zu relativieren.<sup>402</sup> Denn nach diesem Modell hat eine Person schlicht keinen guten Grund, etwas zu tun, wenn die für die Theorie als notwendig erachteten psychischen Zustände kontingenterweise nicht vorhanden sind. Letzteres erscheint problematisch hinsichtlich v.a. der Geltungsansprüche

---

<sup>399</sup> Vgl. Dancy 2000; Bittner 2011.

<sup>400</sup> „Denn manche Handlungen werden ausgeführt, weil etwas für sie zu sprechen scheint, obwohl in Wirklichkeit nichts für sie spricht.“ (Mantel 2009, 260)

<sup>401</sup> Vgl. Smith 1994.

<sup>402</sup> Vgl. Schälike 2002. Diese Position wird – zu Recht oder Unrecht – oft mit Hume in Verbindung gebracht und auch im Zuge nonkognitivistischer Begründungen von Ethik verwendet.

von Aussagen über kluges und moralisches Handeln, die nicht als von solchen kontingenten Umständen abhängig erscheinen.

Sowohl Psychologismus als auch Antipsychologismus greifen auf intuitiv einleuchtende Eigenschaften von Handlungsgründen zurück, die sich aus ihrer funktionalen Doppelrolle als Handlungsrechtfertigung und Handlungserklärung ergeben. Auf Basis der handlungserklärenden Funktion von Gründen ergeben sich folgende *psychologistischen* Argumente:

a) Fakten wie ‚dass es regnet‘ oder ‚dass die Müllabfuhr jeden Dienstag kommt‘ können ein Handeln nicht von sich aus erklären, also das Mitnehmen eines Regenschirms oder das Hinaustragen des Mistkübels. Wenn die Person diese Fakten nicht wahrnimmt oder auf andere Weise kognitiv verarbeitet, so können sie die Handlungen der Person nicht erklären. Die Person braucht demnach einen subjektiven Zugang zu diesen Fakten, damit diese überhaupt eine Rolle spielen können für die Erklärung seines Tuns. Sonst, so der Vorwurf, wird das Handeln aus Gründen erklärt, die der Person überhaupt nicht bewusst waren. Als Reaktion auf diesen Einwand akzeptiert der Antipsychologismus, dass der Glaube des Handelnden eine notwendige Bedingung oder Ermöglichungsbedingung (*enabling condition*) dafür sei, dass er einen motivierenden Grund habe.<sup>403</sup>

b) Reine Gehalte von mentalen Zuständen können ein Handeln ebenso nicht erklären, da diese nicht real und damit kausal wirksam sind. Logische oder propositionale Zusammenhänge verursachen nichts, und damit sind sie auch nicht geeignet für kausale Erklärungen.

c) Die Art der Erklärung einer Handlung müsse unabhängig davon sein, ob die handlungserklärenden Faktoren tatsächlich zutreffen oder nicht (ob sie wahr sind oder nicht). Es gibt aber eine Klasse von Fällen von Handlungen, deren Gründe so beschaffen sind, dass sie die Handlung erklären, obwohl die Gründe auf falschen Überzeugungen des Akteurs basieren, und damit keine Fakten sondern bloße Sachverhalte darstellen. Da Fakten aber definitionsgemäß wahr sind, können demnach nicht alle Erklärungen durch Gründe auf Fakten basieren. Das heißt, dass die *Art* der Handlungserklärung abhängig wäre von der Wahrheit oder Falschheit der erklärenden Faktoren. Letzteres verstößt aber gegen ein plausibles Kriterium für die Adäquatheit von Erklärungen.

Im Zusammenhang mit c) diskutiert Erich Ammereller folgendes Argument, das für eine psychologistische Interpretation spricht:<sup>404</sup>

(P1) Intentionale Handlungen sind Handlungen aus einem Grund

---

<sup>403</sup> Vgl. Ammereller 2005; Mantel 2009, 263.

<sup>404</sup> Vgl. Ammereller 2005, 64f. Vgl. auch die Diskussion bei Dancy 2000.

- (P2) Doppelte Faktivität von Erklärungen: Wenn eine Erklärung ‚q, weil p‘ zutrifft, so sind sowohl p als auch q wahr. Angewendet auf Erklärungen für intentionale Handlungen ergibt sich daraus folgender Unterschied: Jede intentionale Handlung lässt sich nach dem psychologistischen Schema erklären durch die Angabe eines geeigneten Wunsch-Überzeugungs-Paares. Dagegen lassen sich Handlungen nach dem anti-psychologistischen Schema nur dann erklären, wenn der Grund auch tatsächlich besteht, die Überzeugung des Handelnden mithin wahr ist.
- (P3) Die Form der Erklärung einer Handlung ist einheitlich und daher nicht abhängig von der Wahrheit oder Falschheit der Überzeugungen des Handelnden.
- (K) Also: Erklärungen nach dem psychologistischen Schema erfüllen alle drei Adäquatheitsbedingungen für Erklärungen intentionaler Handlungen, Erklärungen nach dem anti-psychologistischen Schema erfüllen diese nicht.

Das psychologistische Schema erkläre nach diesem Argument demnach Handlungen umfassender als das anti-psychologistische, außerdem lassen sich anti-psychologistische Erklärungen immer in psychologistische Erklärungen umwandeln. Die Erklärung ‚X tut Y, weil Z‘ – ‚Hans mäht den Rasen, weil das Gras zu hoch ist‘ – lässt sich umwandeln in ‚X tut Y, weil sie Z will und glaubt, dass Y ein geeignetes Mittel ist, um Z zu erreichen‘ – ‚Hans mäht den Rasen, weil er einen kurzen Rasen haben will und glaubt, dass er dies durch Rasenmähen erreichen kann‘.

Kritisiert wird dieses Argument durch die Aufgabe einer oder mehrerer ihrer Prämissen:<sup>405</sup> a) Nicht alle intentionalen Handlungen sind Handlungen aus einem Grund; b) Handlungserklärungen sind nicht faktiv, d.h. aus der Wahrheit der Erklärung folgt nicht die Wahrheit der Umstände, welche die Handlung erklären; c) Disjunktivismus von Handlungserklärungen: Handlungserklärungen sind ihrer Form nach verschieden, je nachdem, ob die Überzeugung des Akteurs zutrifft oder nicht.

Auf *antipsychologistischer* Seite ergeben sich Argumente gegen psychologistische Konzeptionen auf der Basis der rechtfertigenden Funktion von Handlungsgründen bzw. ihrer Normativität:

a) Rechtfertigung sei eine ideale Beziehung zwischen propositionalen Gehalten, keine reale zwischen psychischen Entitäten. Aus diesen lässt sich nichts ableiten, und zwischen diesen bestehen auch keine Folgerungsbeziehungen. Der Psychologismus identifiziere damit etwas als motivierenden Grund, das als Entität ungeeignet ist, in einer Rechtfertigungsbeziehung zu stehen.<sup>406</sup>

---

<sup>405</sup> Vgl. Ammereller 2005, 74ff.

<sup>406</sup> Vgl. die Unterscheidung bei Mantel zwischen *Zustandspsychologismus* und *Tatsachenpsychologismus*: der Zustandspsychologismus identifiziert psychische Zustände mit motivierenden Gründen, der Tatsachenpsychologismus Tatsachen über das Vorliegen psychischer Zustände. Vgl. Mantel 2009, 261; Ammereller 2005, 70f.

b) Normalerweise spielen die eigenen psychischen Zustände in der praktischen Überlegung von Personen keine explizite Rolle. Die Person denkt über nicht-psychische Tatsachen nach und wie sie auf diese reagieren soll, und nicht über den Zustand der eigenen Psyche und wie sie am besten auf diesen reagieren soll.<sup>407</sup> Diesen Punkt akzeptieren zumeist psychologistische Theorien.

c) Normalerweise spielen für die Handlungsrechtfertigung auch psychische Tatsachen keine Rolle, also Tatsachen, dass bestimmte psychische Zustände vorliegen. Peter entscheidet sich bei einem Pferderennen auf eines der Pferde zu setzen. Der Grund, weshalb er glaubt, dass ein bestimmtes Pferd gewinnt, ist nicht, ‚dass er glaubt, dass dieses Pferd gewinnen wird‘, sondern nur ‚dass es gewinnen wird‘. Denn die Tatsache, dass er etwas glaubt, steigert die Chancen des Pferdes zu gewinnen in keiner Weise und ist deshalb auch kein Grund, an dessen Sieg zu glauben. Manchmal können Fakten über etwas Psychisches allerdings Gründe sein, etwas zu tun. So könnte die Tatsache, gerade besonders aufgeregt zu sein, dagegen sprechen, eine Operation durchzuführen, da die Gefahr besteht, dass man aufgrund der Aufregung zittert und so die Gesundheit des Patienten gefährdet. Jonathan Dancy hat in diesem Zusammenhang eine Adäquatheitsbedingung an eine Theorie praktischer Gründe gestellt: praktische Gründe müssen so beschaffen sein, dass Akteure auch tatsächlich nach ihnen handeln können.<sup>408</sup> Wenn motivierende Gründe also identisch wären mit psychischen Tatsachen, könnten Akteure nur aus Gründen handeln, die keine guten oder rechtfertigenden Gründe sind, oder anders gesagt, der Grund, aus denen jemand handelt, kann die Handlung niemals rechtfertigen.<sup>409</sup>

Wenn zwei philosophische Positionen einerseits jeweils für sich plausibel sind, andererseits jedoch einander ausschließen, gelingt oftmals eine Lösung des Problems nur durch eine Überwindung der stillschweigenden geteilten Voraussetzungen, auf denen die jeweilige Debatte basiert. In diesem Fall vermute ich, dass die realistische Ausgangsposition beider Positionen, sowohl der psychologistischen wie der anti-psychologistischen, eine versöhnende Mittelposition verhindert. Für beide Seiten sind Gründe Fakten in einem realistischen Sinn, d.h. unabhängig von jedem epistemischen Zugriff entweder existierend oder nicht. Dies ist insofern wenig plausibel, da jede Situation eine Unmenge an Fakten beinhaltet, von denen nur ein kleiner Teil

---

<sup>407</sup> Vgl. das sogenannte Spezialfallargument bei Mantel 2009, 263ff.

<sup>408</sup> „This objection amounts to the introduction of a further constraint, which we can call the normative constraint, in addition to the explanatory constraint that we have already seen. This requires that a motivating reason, that in the light of which one acts, must be the sort of thing that is capable of being among the reasons in favour of so acting; it must, in this sense, be possible to act for a good reason. The explanatory constraint held that all normative reasons should be the right sort of thing to contribute to motivation, since that is what they must be if they are to be capable of contributing to the explanation of action in the right sort of way. The normative constraint goes in the other direction, claiming that motivating reasons should be the right sort of thing to be normative reasons.” (Dancy 2000, 103)

<sup>409</sup> Vgl. das sogenannte Idealfallargument bei Mantel 2009, 272ff.

Handlungsgründe für eine Person ausmachen. Gründe, die von niemandem jemals eingesehen werden können, sind ähnlich zu mathematischen Wahrheiten, die niemals jemand beweisen könnte. Eine anti-realistische Interpretation der Objektivität von Gründen könnte hier vielleicht helfen, die starre Opposition aufzubrechen.

Unabhängig davon, wie diese Debatte entschieden wird, möchte ich die Konsequenzen beider Positionen für das Verständnis von Willensschwäche skizzieren. Aus psychologischer Sicht besteht der motivierende Grund, weshalb die Person die willensschwache Handlung getan hat, aus der Tatsache, dass bestimmte psychische Zustände vorlagen. Je nach Auffassung darüber, was die motivierenden Gründe aus psychologischer Sicht genau sind, werden dabei verschiedene psychische Zustände identifiziert bzw. Tatsachen, dass diese vorliegen. Normativ gesehen lag nur eine teilweise Rationalisierung der Handlung vor, da zwar etwas aus Sicht der Person für die Handlung gesprochen hat, aber andere Gründe bestanden, die gegen die Ausführung der Handlung gesprochen haben. Wenn normative und motivierende Gründe voneinander verschieden sind, kann die willensschwache Person motivierende Gründe gehabt haben, in einer Weise zu handeln, ohne über normative Gründe verfügt zu haben, welche die Handlung auch rechtfertigen. Wenn normative Gründe mit motivierenden Gründen identisch sind, ist nicht klar, wie dies möglich sein sollte.

Aus anti-psychologischer Sicht war der motivierende Grund zugleich der normative Grund, der aus Sicht der Person scheinbar für die willensschwache Handlung gesprochen hat. Dieser kann aber nur teilweise für die willensschwache Handlung gesprochen haben, da die Person selbst in ihrer Deliberation über ein besseres Urteil verfügt, dass gegen die willensschwache Handlung spricht. Die anti-psychologische Theorie muss daher annehmen, dass Personen auch aus Gründen handeln können, die nur teilweise für die jeweilige Handlung sprechen, oder davon ausgehen, dass im Fall von Willensschwäche die Person tatsächlich nicht aus Gründen handelt.

### **IV.3.3 Wunsch- und wertebasierte Theorien praktischer Gründe**

Diese Theorien geben spezifische Interpretationen für die jeweiligen psychischen oder nicht-psychischen Fakten an, welche Gründe sein sollen. Für wunschbasierte Theorien praktischer Gründe besteht der Grund, eine Handlung zu tun, aus einer psychologischen Tatsache einer Person, genauer dem Faktum, dass die Person einen bestimmten Wunsch nach etwas hegt, oder dem Faktum, dass die Person bestimmte Wunsch-Überzeugungs-Paare hat.<sup>410</sup> Der Grund, warum Christian den Fernseher einschaltet, ist, weil er einen Wunsch danach hat, die Übertragung des WM-Finales zu sehen. Daniel hat keinen entsprechenden Wunsch, deshalb hat er keinen Grund, den Fernseher einzuschalten.

---

<sup>410</sup> Vgl. Mark Schroeder 2007 für eine Verteidigung einer humeanischen Theorie praktischer Gründe.

„On a widely accepted story of human agency, *all* reasons for acting, intending, and desiring are provided by the fact that the agent wants something or would want it under certain conditions. My reason for going to the store, for example, is provided by the fact that I want to buy some ice cream, and my reason for wanting to buy some ice cream is provided by the fact that I want to eat some. According to these ‘desire-based’ accounts, all practical reasons are grounded in the present desires of the agent; justification has its source in the fact that *I do or would want it.*” (Chang 2004, 56)

Diese Sicht von normativen Gründen fügt sich recht gut in kausale Handlungstheorien ein, da diese darauf beruhen, bestimmte Entitäten als Ursachen für Handlungen zu identifizieren, z.B. mentale Ereignisse (Wunsch-Überzeugungs-Paare), und über diese Ursachen den Handlungsbegriff zu explizieren. Sie passt auch gut zu einem Modell praktischer Rationalität als instrumentelle Vernunft: der Wunsch, etwas zu tun, ist abgeleitet von einem Wunsch nach etwas anderen. Die Ableitung geschieht über die Überzeugung, die Handlung sei Mittel für dieses andere, in Gemeinschaft mit dem Prinzip instrumenteller Rationalität, die Mittel zu erstreben, die für die Erreichung eines Zieles hinreichend sind. In der entscheidungstheoretischen Interpretation entsprechen den beiden psychischen Faktoren die Präferenzstruktur und die Glaubenszustände hinsichtlich der Wahrscheinlichkeiten für das Eintreten der präferierten Zustände.

Welchen Grund hat nach diesem Modell die willensschwache Person, ihre Handlung statt der bevorzugten zu tun? Sie braucht ein entsprechendes Wunsch-Überzeugungs-Paar, welches die willensschwache Handlung verursacht und diese rechtfertigt. Die Rechtfertigung sollte dem Standard instrumenteller Vernunft genügen. Nun urteilt dieselbe Person, dass für eine von zwei alternativen Handlungen, alles in allem betrachtet, bessere Gründe sprechen als für die andere. Nach Davidson besteht das Problem nun darin, dass die normative Stärke des Grundes, aus dem die Person handelt, nicht mit der kausalen korreliert: der Grund, aus dem die Person handelt, ist überhaupt kein Grund für die Handlung oder zumindest schwächer als der Grund für die alternative Handlung. Es braucht hier eine Erklärung, weshalb sich der schwächere Grund manchmal durchsetzen kann.

Für wertebasierte Theorien praktischer Gründe sind Gründe, etwas zu tun, Fakten über die Werte, die durch eine Handlung realisiert werden.

„On these ‘value-based’ accounts, reasons for acting, intending, and desiring are provided by facts about the value of something, where being valuable is not simply a matter of being desired. [...] Value-based views ground all practical reasons in evaluative facts or the facts that subvene them; justification has its source not in the fact that one wants something but in facts about what one wants.” (Chang 2004, 56f.)

„This approach, the classical approach, it may be called, can be characterized as holding that the central type of human action is intentional action; that intentional action is action for a reason; and that reasons are facts in virtue of which those actions are good in some respect and to some degree.” (Raz 1999, 22f.)

Gemäß dieser Art von Theorie wäre der Grund der Person für die willensschwache Handlung eine Eigenschaft der Handlung, welche die Handlung wertvoll macht. Aus dieser Theorie ist nicht

ersichtlich, wie eine Person einen Grund haben kann für ihre willensschwache Handlung, da sie selbst diese als nicht wertvoll im Vergleich zu alternativen Handlungen ansieht. Innerhalb dieser Theorien liegt es deshalb nahe, die Existenz inkommensurabler Werte anzunehmen, zwischen denen die willensschwache Person entscheidet.<sup>411</sup> Die willensschwache Handlung ist deshalb begründet, da sie einen Wert realisiert, der nicht kommensurabel ist zu dem Wert der als besser beurteilten Handlung. Das Problem ist dabei nicht nur, die Inkommensurabilitätsthese verständlich zu machen, ohne dass sie ad hoc klingt, sondern auch daran festzuhalten, dass etwas mit der willensschwachen Handlung nicht stimmt. Denn wenn die Inkommensurabilität auch eine Inkomparabilität beinhaltet, wird unklar, wie die Person zu einem besseren Urteil gelangt. Die Kehrseite davon wäre, dass die willensschwache Handlung nach dieser Vorstellung keinen normativen Mangel aufweist, da alle zueinander inkommensurablen Handlungen gleich gut wären.

Auch in dieser Frage bin ich der Meinung, dass eine Mittelposition eine vernünftige Alternative zu den starren Extrempositionen darstellen kann. Eine der Voraussetzungen der Diskussion ist, dass Gründe *entweder* auf Werten *oder* auf Wünschen basieren, jedoch nicht auf beiden zugleich. Je nach Bereich sind wir im Alltag jedoch bereit, verschiedene Quellen von Gründen als Rechtfertigungen zu akzeptieren. Wenn jemandem Schokolade nicht schmeckt, so hat diese Person einfach keinen Grund, Schokolade zu essen, während ein bestimmtes Maß an moralischer Motivation als Teil dessen angesehen wird, das jeder jedem schuldet. Die Schwierigkeit besteht darin, diese Quellen von Gründen plausibel voneinander abzugrenzen und deren relative Eigenständigkeit zu begründen.

#### **IV.4 Sind willensschwache Handlungen absichtlich?**

Was sind Bedingungen für die Charakterisierung einer Handlung als ‚intentional‘? Dazu stelle ich fest, dass unter Absichtlichkeit von Handlungen, deren Intentionalität, Verschiedenes verstanden wird. In der philosophischen Analyse fungiert der Begriff 1) als Kriterium für die Zuschreibung von Verantwortung für Handlungen, 2) als Kriterium dafür, ob überhaupt eine Handlung vorliegt, 3) als Identitätskriterium für Handlungen, also wann eine Handlung identisch ist mit einer anderen, und 4) als Korrektheitskriterium, also als Antwort darauf, wann jemand eine Handlung korrekt ausgeführt hat.

1) Mit der Absichtlichkeit als Verantwortungskriterium beschäftige ich mich in Kap. V. Vorweg denke ich, dass Intentionalität weder hinreichend noch notwendig ist, um Verantwortung zu-

---

<sup>411</sup> Vgl. die Diskussion in Kap. III.4.

zuschreiben. Nicht für alles, was wir intentional tun, sind wir verantwortlich, und nicht alles, wofür wir verantwortlich sind, ist etwas, was wir intentional tun.

2) Donald Davidson benutzt die Intentionalität von Handlungen als Kriterium dafür, ob eine Handlung vorliegt oder nicht. Eine Handlung ist eine Tätigkeit einer Person, welche unter zumindest einer Beschreibung intentional ist.

„[A] man is the agent of an act if what he does can be described under an aspect that makes it intentional.“  
(Davidson 1971, 46)

„[A] person is the agent of an event if and only if there is a description of what he did that makes true a sentence that says he did it intentionally.“ (Davidson 1971, 46)

Die Intentionalität von Handlungen ist in dieser Theorie keine Eigenschaft der Handlung selbst, sondern eine Relation, die zwischen Beschreibungen von Handlungen und der Aussage, die Handlung sei intentional, besteht.<sup>412</sup> Damit eine Beschreibung einer Handlung die Handlung als intentional erweisen kann, muss sie nach Davidsons Theorie einerseits zeigen, wie sich primäre Gründe konstruieren lassen, welche die Handlung unter dieser Beschreibung als rational aus der Sicht der Person erweisen, und diese Gründe müssen außerdem in einer Kausalrelation zu der Handlung stehen. Das Problem abweichender Kausalketten hat zur Folge, dass diese beiden Bedingungen nicht hinreichend dafür sind, ob etwas eine Handlung ist, d.h. intentional unter einer Beschreibung. Dies deshalb, da beide Relationen bestehen können, ohne miteinander zu harmonieren: das Resultat ist eingetreten, und der Akteur hatte die Absicht, dieses Resultat herbeizuführen, dennoch hat er nicht absichtlich gehandelt bzw. das Resultat absichtlich herbeigeführt. Die korrespondierende Frage zu Willensschwäche ist, ob willensschwache Tätigkeiten überhaupt Handlungen sind. Gemessen an der Diskussion bisher hängt dies daran, ob man eine Theorie der Willensschwäche vertritt, bei der die Person zumindest etwas absichtlich tut, wenn sie sich willensschwach verhält. Die Theorien, die dies verneinen, setzen dabei Theorien über den Zusammenhang von evaluativen Urteilen und Intentionen voraus, die nicht zwingend sind. Unter 4) zeige ich, dass bei einem differenzierten Verständnis der Intentionalität von Handlungen willensschwache Handlungen zumindest als zielgerichtete und begründete Handlungen aufgefasst werden sollten, auch wenn sie von einem Idealbild des Handelns aus Gründen abweichen.

3) Die Absichtlichkeit oder Intentionalität als Identitätskriterium von Handlungen betrifft die Frage, *welche* Handlung von einer Person überhaupt ausgeführt wurde und welche der Tätigkeiten einer Person miteinander identisch sind. Anscombe und Davidson haben in der Frage der Identität von Handlungen dabei einen *grobkörnigen* Ansatz vertreten, während Handlungstheoretiker wie Alvin Goldman einen *feinkörnigen* Ansatz vorgebracht haben. Nach einer

---

<sup>412</sup> Vgl. Löhner 2003, 31–58.

grobkörnigen Theorie gibt es zu jeder Handlung verschiedene Beschreibungen, für den feinkörnigen Ansatz ist jede Beschreibung einer Handlung mit einem entsprechenden Handlungstyp verbunden und beschreibt damit, falls sie zutrifft, eine Handlung. Ein Beispiel soll den Unterschied zwischen den beiden Theorietypen veranschaulichen:

„I flip the switch, turn on the light, and illuminate the room. Unbeknowst to me I also alert a prowler to the fact that I am home. Here I need not have done four things, but only one, of which four descriptions have been given.” (Davidson 1963, 4)

Nach Davidson und Anscombe sind das Anknipsen des Schalters, das Anschalten des Lichts, das Beleuchten des Raumes und das Alarmieren des Einbrechers verschiedene Beschreibungen einer einzigen Handlung. Nach Goldman führt die Person vier verschiedene Handlungen aus, welche jedoch systematisch miteinander zusammenhängen.<sup>413</sup> Im Unterschied zu Anscombe benutzt Davidson die Intentionalität von Handlungen nicht als deren Identitätskriterium. Bei Anscombe bezeichnen zwar verschiedene Beschreibungen von Handlungen jeweils dieselbe Handlung, doch stehen die diesen Beschreibungen entsprechenden Absichten in logischen Verhältnissen zueinander, die asymmetrisch sind und sich letztlich zu einer finalen Absicht zusammenfassen lassen.

„Demnach haben wir *eine* Handlung mit vier Beschreibungen vor uns,<sup>414</sup> deren jede von umfassenderen Umständen abhängt und die zur jeweils nächsten im Verhältnis von Mittel- und Zweckbeschreibung stehen. Das wiederum bedeutet, daß wir genausogut von *vier* entsprechenden Absichten sprechen können wie von *einer* entsprechenden Absicht, nämlich dem letzten Glied, das wir in die Reihe eingefügt haben.“ (Anscombe 2011, 75)

Für Davidson sind dagegen Aussagen über Handlungen im Grunde Existenzquantifikationen über Ereignisse.<sup>415</sup> Partikuläre Ereignisse übernehmen daher die Rolle als Identitätskriterium für Handlungen: eine Handlung ist identisch mit einer anderen, wenn ihre Beschreibung auf dasselbe partikuläre Ereignis referiert bzw. dasselbe Ereignis bezeichnet wie die andere Handlung. Für die Identität von Handlungen spielt demnach deren Intentionalität keine Rolle. Das Problem der Identität von Handlungen verlagert sich dadurch in ein Problem der Identität von Ereignissen. Davidson hat dafür zuerst ein kausales – Ereignisse sind dann identisch, wenn sie dieselben Ursachen und Wirkungen haben –, danach ein raum-zeitliches Kriterium – Ereignisse sind identisch, wenn sie dieselbe Raum-Zeitstelle einnehmen – vorgeschlagen.<sup>416</sup>

---

<sup>413</sup> „Since an act-token is the exemplifying of a property by an agent at a time, it is natural so to individuate act-tokens that *two act-tokens are identical if and only if they involve the same agent, the same property, and the same time.*” (Goldman 1970, 10)

<sup>414</sup> Anscombe bezieht sich dabei auf ihr Beispiel mit einem Mann, der seinen Arm bewegt, eine Wasserpumpe betätigt, den Wasservorrat eines Hauses auffüllt und die dortigen Bewohner vergiftet, vgl. Anscombe 1963, §23,37ff.

<sup>415</sup> Dies zeigt sich auch formal in seiner logischen Analyse von Handlungssätzen in Davidson 1967.

<sup>416</sup> Vgl. Davidson 1969b; 1985b.

Die Gleichwertigkeit verschiedener Beschreibungen zueinander – jede Beschreibung, die eine Handlung eindeutig bezeichnet, ist gleich gut – wird von Davidson in weiterer Folge durch eine Privilegierung von Beschreibungen von Körperbewegungen unterlaufen. Dabei greift er auf das Konzept von Basisakten zurück, wie es von Arthur C. Danto in die Diskussion eingebracht wurde.<sup>417</sup> Während Danto noch zwischen Basishandlungen und komplexen Handlungen unterscheidet, die von Basishandlungen verursacht werden, sind Basishandlungen nach Davidson auch bereits die Menge aller Handlungen, die eine Person überhaupt vollziehen kann. Handlungen, die vorgeblich durch Basishandlungen verursacht werden, sind lediglich Beschreibungen von Basishandlungen unter einem anderen Aspekt. Letzteres betrifft den sogenannten *Akkordeon-Effekt*,<sup>418</sup> die Möglichkeit, Handlungen über ihre Effekte und Folgen neu zu beschreiben:

„We must conclude, perhaps with a shock of surprise, that our primitive actions, the ones we do not by doing something else, mere movements of the body – these are all the actions there are. We never do more than move our bodies: the rest is up to nature.” (Davidson 1971, 59)

Frederick Stoutland kritisiert diese Privilegierung von Beschreibungen von Handlungen durch rein physische Beschreibungen von Körperbewegungen gegenüber Beschreibungen mithilfe psychischer Begriffe.<sup>419</sup>

„What does it mean to say that action consists in bodily movements? The idea is that only bodily movement descriptions – or more generally, physical descriptions – identify the descriptum for action descriptions *all by itself, by reference to nothing else*. Other descriptions – in particular psychological descriptions – describe the behavior *by reference to* something other than the behavior – for example, by reference to its causes or effects.” (Stoutland 1985, 55)

Gegen diese Ansicht wendet Stoutland ein, dass es psychologische Beschreibungen sind, welche es uns ermöglichen, etwas als *eine* Handlung aufzufassen, und nicht Beschreibungen von Körperbewegungen. Bei komplexen Handlungen wie dem Verfassen einer Email kann man sich das folgendermaßen klarmachen: die unzähligen Bewegungen der Finger und des gesamten Körpers bekommen ihre Einheit durch die psychologische Beschreibung als ‚Verfassen einer Email‘. Worin sollte sich sonst das Schreiben einer Email vom Schreiben von zwei oder mehreren Emails unterscheiden? Die Körperbewegungen wären in beiden Fällen dieselben.<sup>420</sup>

„To identify *an* action, which is such that *it* has a physical description, assumes a description which contains a count term which yields one action of that kind. Such a description will be psychological. Indeed normally it will be a description under which the behavior is intentional.” (Stoutland 1985, 56)

„Once we have identified an action, we can construct a physical description which describes it. But the physical description will in the usual cases include count terms which count several descripta, not one. The

---

<sup>417</sup> vgl. Danto 1965.

<sup>418</sup> Vgl. Davidson 1971, 53–61.

<sup>419</sup> Vgl. Stoutland 1985, 55ff.

<sup>420</sup> Für kausale Handlungstheorien unterscheiden sich diese Handlungen durch ihre Ursachen. Doch wie lassen sich wiederum die verschiedenen Ursachen und Wirkungen voneinander unterscheiden?

action will be *an* action only when counted in terms of psychological descriptions. As far as action is concerned, therefore, physical descriptions are secondary to psychological ones.” (Stoutland 1985, 56)

Einen alternativen Zugang zur Identität von Handlungen beschreibt Guido Löhrer in seinem Buch *Praktisches Wissen* (2003). Aufbauend auf intuitionistischer Logik und konstruktiver Typentheorie in der Version von Per Martin-Löf sind Handlungen *Zwecke*, die durch ihre *Mittel* individuiert werden.

„Ein Zweck ist der Typ der ihn realisierenden Mittel.“ (Löhrer 2003, 208)

Die Kriterien für die Identität von Handlungen beziehen sich daher auf die Menge an Mittel, die einen Zweck realisieren.

„Ist bekannt, daß A ein Zweck und B ein Zweck ist, so bedeutet  $A = B : \text{Zweck}$  (i) daß, wenn a ein Mittel ist, das A herbeiführt, a auch ein Mittel ist, daß B herbeiführt, und umgekehrt, und (ii) daß, wenn a und b die gleichen Mittel im Hinblick auf die Verwirklichung von A (gleiche Objekte des Typs A) sind, a und b auch die gleichen Mittel im Hinblick auf die Verwirklichung von B sind [...].“ (Löhrer 2003, 335)

Hier ist leider nicht der geeignete Ort, um näher auf die logische Eleganz und Stringenz dieses Ansatzes einzugehen. Es soll nur zeigen, dass teleologische Handlungstheorien, mit geeigneten logischen Mitteln ausgestattet, über zumindest gleichwertige Methoden verfügen, sinnvoll über die Identität von Handlungen zu sprechen wie ihre extensionalistisch geprägten kausalen Alternativen.

Die Frage, wann zwei willensschwache Handlungen identisch sind, hieß demnach in einem extensionalen Ansatz nach Davidson, dass beide Beschreibungen auf dasselbe Ereignis, welches als Körperbewegung eindeutig beschrieben werden kann, verweisen. Nach einem teleologischen Ansatz, ob die beiden Handlungen durch dieselben Mittel individuiert werden, die sie herbeiführen. Beide Urteile sind möglich, solange feststeht, dass die Person überhaupt handelt, mithin etwas absichtlich tut.

4) Ich schlage vor, den Begriff der ‚Intentionalität‘ im Sinne der Korrektheit für Tätigkeiten oder Verhalten auf verschiedenen Niveaus zu gebrauchen. Korrelativ dazu sind die Kriterien, wann eine Handlung oder Tätigkeit als absichtlich zu charakterisieren ist, verschieden anspruchsvoll. Ich möchte zwischen zumindest drei Stufen ‚intentionalen Handelns‘ unterscheiden, die aufeinander aufbauen: ein Wesen, das die Fähigkeit zu einer höheren Form intentionalen Handelns und Verhaltens besitzt, verfügt normalerweise auch über die darunterliegenden Formen.

Handlungstheorien unterscheiden sich darin, ob sie intentionales Handeln eher ‚bottom-up‘ oder ‚top-down‘ analysieren. Während sich die einen an natürlicher, auch bei anderen Lebewesen vorfindlicher Intentionalität orientieren, nehmen die anderen das vollstrukturierte, durch Überlegungsprozesse angestoßene Handeln aus Gründen als Vorbild für die Analyse. Für

manche Positionen schließen sich die beiden Betrachtungsweisen (top-down und bottom-up) einander strikt aus. Natürliche Prozesse als funktional oder zweckmäßig zu betrachten sei verschieden von der Intentionalität menschlicher Handlungen. Um Intentionen, Überzeugungen, Wünsche etc. zu haben muss ein Subjekt über bestimmte reflexive Fähigkeiten verfügen.

„To place my own cards on the table, I take seriously the idea that the beliefs and intentions that figure in our thinking about each other are inextricable from reflective capacities, including those necessary for appreciating the normative commitments incurred by beliefs and intentions. I also take seriously the idea that using words meaningfully implicates reflective capacities, including those necessary for appreciating the normative commitments incurred in virtue of meaning something by a word.” (Millar 2004, x)

Demgegenüber sehe ich Ansätze als vielversprechender an, welche die Zuschreibung von mentalen Prädikaten auch an Lebewesen gestattet, die über keine Sprache und keine offensichtlichen reflexiven Fähigkeiten verfügen.

Folgende Ebenen von Intentionalität möchte ich voneinander unterscheiden. Zentral für die Zuschreibung von Intentionalität von Verhalten sind dabei unter anderem das Vorliegen von Kontrolle (Kontrolle über Handlungsbeginn und Handlungsverlauf, Reaktion auf Hindernisse, Handlungsunterbrechung und -fortsetzung), Zielgerichtetheit (Orientierung an einen Zielzustand) und Evaluation (Überprüfung des Resultates und Überwachung des Handlungsverlaufes). Diese drei Faktoren werden in unterschiedlichem Maß und auf unterschiedliche Weise auf den drei Stufen realisiert.

a) *Basale Intentionalität* (zielgerichtetes Verhalten): Dazu gehört die prinzipielle Ausrichtung des Verhaltens an einen Zielzustand, die Kontrolle des Verhaltens durch ständige Evaluation des Lebewesens bzw. der Bewegung anhand dieses Zustandes sowie Reaktion auf Veränderungen. Dazu gehören auch Anpassungsleistungen des Tätigkeitsverlaufes an sich verändernde Umweltzustände. Auf diese Weise lassen sich auch bei Tieren intentionale von nicht-intentionalen Bewegungen unterscheiden:

„Consider the difference between what goes on when a spider moves its legs in making its way along the ground, and what goes on when its legs move in similar patterns and with similar effect because they are manipulated by a boy who has managed to tie strings to them. In the first case the movements are not simply purposive, as the spider’s digestive processes doubtless are. They are also attributable to the spider, who makes them. In the second case the same movements occur but they are not made by the spider, to whom they merely happen.” (Frankfurt 1978, 78)

Diese natürliche und animalische Form von intentionaler Bewegung und intentionalem Verhalten ist meiner Ansicht nach nicht abhängig von propositionalen Einstellungen. Die Zielgerichtetheit kann in der Funktionalität der jeweiligen Bewegungen liegen, welche sich als Resultat evolutionärer Prozesse verstehen lässt. So lässt sich das Kümmern der Entenmutter um ihr Junges als zielgerichteter Prozess verstehen, der nicht in mentalen Einstellungen oder deren

propositionalen Gehalten begründet ist, sondern in der Zweckmäßigkeit in evolutionärem Sinne.<sup>421</sup>

Wichtig ist, dass auch bei diesen Prozessen und zielgerichteten Bewegungen unterschieden werden kann, ob ein Resultat erreicht wurde oder nicht, und ob dieses absichtlich erreicht wurde, d.h. hier durch den Prozess selbst und auf die richtige Weise, oder nicht. Ein Indiz für die Richtigkeit dieser Annahme ist, dass auch hier abweichende Kausalketten die Zuschreibung von Intentionalität zu ansonsten funktionierenden zielgerichteten Prozessen untergraben. Die Grundidee ist, dass es auch für animalische Bewegungen und animalisches Verhalten Kriterien gibt, wann diese korrekt vorliegen und *erfolgreich* umgesetzt wurden. Auch Tiere machen Fehler in der Ausführung ihres Verhaltens und stellen sich besser oder schlechter an. Natürlich beschreiben Tiere mangels einer Sprache ihre eigenen Tätigkeiten nicht, sie können daher diese auch nicht unter bestimmten Beschreibungen intendieren.

b) *präreflexive Intentionalität* (begründetes Handeln): Intentionale mentale Einstellungen, welche die Handlung kontrollieren und deren propositionaler Gehalt die Handlung rechtfertigen. Wichtig ist, dass der Akteur über keine Repräsentation dieser Rechtfertigungsbeziehung selbst verfügt bzw. dass der Person die Rechtfertigungsbeziehung nicht thematisch oder thematisch bewusst ist. Das Handeln ist demnach sensitiv für mentale Einstellungen des Akteurs, aber die Rechtfertigungsbeziehung zwischen gegebenen Fakten, ihnen angemessenen Reaktionen und korrekten mentalen Einstellungen wird nicht repräsentiert. Ein Kind, das merkt, dass es hungrig ist, und etwas Kuchen vom Tisch nascht, handelt intentional, auch wenn es sich selbst sein Handeln nicht als Begründetes repräsentiert. Dies entspricht einem externalistischen Konzept der Rechtfertigung in der Epistemologie.

c) *reflektierte Intentionalität* (Handeln aus Gründen): Interne Repräsentation bzw. ein reflexives Bewusstsein der Rechtfertigungsbeziehung zwischen den propositionalen Gehalten der intentionalen Einstellungen und der Handlung. Die Person handelt nicht nur begründet, sondern kann sich selbst die Rechtfertigungsbeziehung zwischen ihren Gründen und Handlungen einsichtig machen. Das Bestehen dieser Rechtfertigungsbeziehung kann selbst zu einem Faktor innerhalb der praktischen Überlegung der Person werden, ob sie die Handlung ausführen soll oder nicht. Wie diese Beziehung der Rechtfertigung genau beschaffen ist, ist Teil der Diskussion über die Natur praktischer Gründe. Bei Davidson sind die einfachsten Einheiten dieser Beziehung Wunsch-Überzeugungs-Paare und Handlungen unter bestimmten Beschreibungen. In teleologischen Modellen sind es Werte, verkörpert durch Handlungsziele, und Handlungen unter bestimmten Beschreibungen.

---

<sup>421</sup> Vgl. die Untersuchung von Dretske 1999 für natürliche Formen von Zielgerichtetheit

Unter welche dieser drei Stufen fällt nun die Intentionalität willensschwacher Handlungen? Willensschwache Handlungen erfüllen den Grad von Intentionalität der Stufe a. Sie sind kontrolliert und zielgerichtet. Die Torte, die eine willensschwache Person isst, fällt nicht zufälligerweise in ihren Mund. Wie sieht es mit den Stufen b und c aus? Für sich genommen ist die Handlung lediglich begründet, aber nicht aus Gründen vollzogen. Dies deshalb, da die Person selbst urteilt, dass die Rechtfertigungsbeziehung zwischen ihren Gründen und der Handlung nicht besteht, sie daher die Handlung nicht *deswegen*, also wegen dem Bestehen der Rechtfertigungsbeziehung, vollziehen könnte. Die willensschwache Handlung ist aber begründet, da unter anderen Umständen die propositionalen Gehalte der mentalen Einstellungen der Person ihr Handeln durchaus rechtfertigen könnten. Die Lust auf eine süße Torte begründet gemeinsam mit entsprechenden Überzeugungen das Essen der Torte. Oder aus Sicht einer antipsychologischen Theorie von Handlungsgründen: Das Essen der Torte ist etwas Wertvolles, und kann unter anderen Umständen ein geeignetes Ziel einer Handlung sein. Willensschwäche ist deswegen kein reiner Fall von Handlungen der Stufe b, da die Prozesse, die normalerweise zu Handlungen der Stufe c führen, misslingen. Nur in *Relation* zu diesen Prozessen kann die willensschwache Handlung überhaupt als willensschwach charakterisiert werden, ansonsten wäre sie einfach nur eine Handlung des Typs b oder c. Die Handlung, welche in diesem Falle scheitert, ist die Handlung der praktischen Deliberation. Deren Ziel ist es, das daran anschließende Handeln über die besten Gründe zu bestimmen. Die willensschwache Person scheitert daran, dieses Ziel zu verwirklichen. In kausaler Terminologie rationalisieren die mentalen Einstellungen nicht die Beschreibung der willensschwachen Handlung als überlegt. In teleologischer Terminologie gehört die willensschwache Handlung nicht zum Typ überlegten Handelns. In Analogie gleicht es einem Beweisversuch, der nicht funktioniert. Dennoch gehört die willensschwache Handlung zu einem Typ, da die Person etwas intentional tut, und, falls erfolgreich, die nötigen Mittel in korrekter Weise einsetzt um die Handlung zu verwirklichen.

Sind diese Kriterien für eine basale Unterteilung von Handlungen in verschiedene Klassen anhand ihrer Intentionalität nicht nur notwendig, sondern auch hinreichend? Ist eine Handlung absichtlich, wenn sie in eine der Kategorien fällt? Das Problem abweichender Kausalketten zeigt, dass die Kriterien nicht hinreichend sein können. Bereits bei Handlungen der Stufe a) kann es zu Abweichungen kommen, welche die Intentionalität der Handlung (oder zielgerichteten Bewegung) untergräbt. Ein Beispiel: Eine Katze verfolgt eine Maus. Diese Bewegung ist zielgerichtet und steht unter der Kontrolle der Katze. Ändert die Maus ihre Richtung, so ändert die Katze die ihre. Ziel der Bewegung der Katze ist, die Maus zu fangen. Der natürliche Prozess des Jagens ist evolutionär gesehen geeignet für die Beschaffung von Beute, d.h. es ist ein Faktor, der sich positiv auf die generelle Kapazität zur Reproduktion der Katze und damit ihrer

Nachkommen auswirkt. Nun kann es sein, dass dieser zielgerichtete Prozess sein Ziel erreicht, die Katze die Maus fängt, und das Jagen der Katze kausal verursacht, dass sie die Maus fängt, die Kausalkette aber so abwegig ist, dass die Katze für uns als Beobachter die Maus nicht absichtlich gefangen hat. Die Katze könne z.B. beim Laufen eine Falle auslösen, die von einem besorgten Hausbewohner aufgestellt wurde, welche die Maus durch einen elektrischen Schlag tötet. Das Jagen der Maus verursacht kausal das Ziel, die Beute zu fangen. Dieses singuläre Kausalurteil fällt nicht nur unter allgemeine Regularitäten<sup>422</sup> wie ‚Auf Mäusejagden von Katzen folgt mit Wahrscheinlichkeit P der Tod der Maus‘ sondern stützt auch kontrafaktische Aussagen wie ‚Hätte die Katze die Maus nicht gejagt, so wäre die Maus nicht gestorben‘. Dennoch hat die Katze die Maus nicht zielgerichtet (‚absichtlich‘ oder korrekt) gefangen, da der natürliche Prozess ihres Jagdverhaltens kein Produkt des Zusammenspiels von elektrischen Fallen und zielgerichteten Bewegungen ist. Es ist purer Zufall, dass in diesem Falle die Katze die Maus durch ihr Jagen erlegen konnte.<sup>423</sup> Wichtig an den Beispielen zu abweichenden Kausalketten ist immer zu beachten, dass Kausalketten von sich aus nicht abweichend sind: sie verlaufen so, wie sie eben verlaufen, eine ist so gut wie die andere. Die Abweichung ist eine Zuschreibung, die nur auf Basis einer Vorstellung von Normalität funktionieren kann, in diesem Falle der Normalität aus der Sicht von Beobachtern, die zwischen zielgerichteten und nicht-zielgerichteten Handlungen unterscheiden wollen. Wenn es nun nicht an dem Vorliegen bestimmter mentaler Zustände alleine liegt, die eine Handlung zu einer absichtlichen machen, woran dann? Es liegt nahe, statt individueller Zustände auf ein geteiltes Wissen Bezug zu nehmen, das Akteure, die zugleich Wissenssubjekte sind, *teilen* und anhand dessen sie entscheiden können, ob bestimmte Tätigkeiten absichtlich sind oder nicht.<sup>424</sup> Dieses Wissen ist ein Wissen darüber, *wie man Dinge tut*, also welche Schritte normalerweise dafür nötig sind, um bestimmte Ziele zu erreichen. Eine Handlung wäre nach dieser Vorstellung eine Zuordnung eines Zieles zu geeigneten Mitteln.<sup>425</sup> Dabei ist dieses Wissen als dynamisch anzusehen, d.h. erfolgreiche Strategien, Ziele durch neue oder andere Mittel zu erreichen, können dieses Wissen verändern, ebenso wie ständige Erfolglosigkeit oder Überprüfungen alter Strategien, die nicht mehr funktionieren, diese aus dem Wissenskanon

---

<sup>422</sup> Das Beispiel ist natürlich so keine wissenschaftlich ernst zu nehmende Regularität, schon aufgrund der vagen Begriffe. Es soll nur hier nur dazu dienen darauf hinweisen, dass Regularitäten in irgendeiner Art angenommen werden, wenn Jagdverhalten und Nahrungszugang positiv zueinander korreliert sind.

<sup>423</sup> Wem dies noch nicht zufällig genug ist, der kann die Kausalkette natürlich beliebig um weitere Zufälle erweitern.

<sup>424</sup> Ich stütze mich hier weitgehend auf Löhner 2003.

<sup>425</sup> „If the end of an action is supposed to be an aim which can be attained by certain means and if the volition of means is given by the volition of an end, then it seems appropriate to understand the end as a type, namely as the type of those means by which the end can be realized. Thus we obtain an end-as-type-principle. If something is an end, then the act realizing this end is an object of the type of this end, namely an object of the type of the means realizing this end. This type is not empty [...] if there is at least one means falling within its scope, that is, that a person who desires this end is able (and willing) to perform an action being an apt means to realize this end. To deliberate is to search for an object of the type in question.” (Löhner 2005, 17)

herausfallen lassen. Guido Löhrer repräsentiert in seiner konstruktiven Handlungstheorie daher Handlungen im Handlungswissen von Agenten.

„Konstruktiv ist die Handlungstheorie in dem Sinne, dass sie Handlungen nicht auf die Welt oder auf ein System möglicher Welten abbildet, sondern im Handlungswissen ihrer Agenten repräsentiert. Dieses für das Handeln und seine Darstellung relevante Wissen nenne ich praktisches Wissen. Handlungen sind durch praktisches Wissen möglich, und jede Handlung führt zu einer Expansion, Revision oder gegebenenfalls auch zu einer Kontraktion dieses Handlungswissens.“ (Löhrer 2003, 13)

Zentral für die hier geführte Diskussion um die Absichtlichkeit von Handlungen ist dabei der ins Spiel gebrachte Begriff der Korrektheit.<sup>426</sup>

„Der Begriff der Korrektheit ist für eine konstruktive Theorie menschlichen Handelns grundlegend. Ohne ihn könnte sie nicht erklären, was eine Handlung ist. Ohne ihn wären Handlungen für sie nicht individuiert und könnten nicht identifiziert werden. Diese Aufgaben involvieren allesamt den Begriff einer Handlungsrichtigkeit als Korrektheit der Mittel zu einem Handlungszweck. Mittel sind die Verfahren, die geeignet und angemessen sind, einen Zweck zu verwirklichen, und durch deren Gebrauch ein Zweck im Fall gelingenden Handelns so verwirklicht wird, dass der Agent behaupten dürfte, er habe ihn verwirklicht. Damit eine Handlung diese oder jene Handlung ist, müssen die für diese Handlung erforderlichen Verfahren eingesetzt werden. Nicht der bloße Erfolg, sondern die Mittel individuierten eine Handlung.“ (Löhrer 2003, 401)

Löhrer unterscheidet zwischen *korrekten* und *inkorrekten* Handlungen.<sup>427</sup>

a) Korrekte Handlungen: Diese unterteilen sich in solche, bei denen *kanonische*, und solche, bei denen *nicht-kanonische* Verfahren eingesetzt werden. Kanonisch sind Verfahren, die in der jeweiligen Gemeinschaft erprobt sind. Nicht-kanonische Verfahren sind Ausführungsvarianten und Vereinfachungen kanonischer Verfahren.

b) Inkorrekte Handlungen: Diese sind entweder *blind*, d.h. aufs Geratewohl probierte Handlungen, bei denen die Person einfach hofft, dass sich das erwünschte Ziel einstellt, ohne aber über das relevante Handlungswissen zu verfügen oder es einzubringen, oder *irrtümlich*, bei denen die Person die falschen Mittel für den gegebenen Zweck einsetzt.<sup>428</sup>

In Anbetracht der vorigen Unterscheidung verschiedener Stufen von intentionalem Handeln und Verhalten stellt sich die Frage, ob der Ansatz, Handlungen im gemeinschaftlich geteilten praktischen Wissen von Akteuren zu verorten, dazu nicht im Widerspruch steht. Insbesondere dann, wenn tierisches Verhalten in die Kategorie bloß zielgerichteten Verhaltens fallen soll und

---

<sup>426</sup> Vgl. auch „Handlungen, ihre Individuierung und Identifizierung, setzen den Begriff der Korrektheit voraus. Ohne ihn lassen sich Handlungen nicht als Handlungen bestimmen. Korrektheit gibt es nicht privat. Aus den genannten Gründen setzt sie eine Sozialität voraus, in der „korrekt“ und „inkorrekt“ nicht allein bestimmbar und unterscheidbar werden, sondern in der diese Bestimmung und Unterscheidung mittels positiver und negativer Sanktionen auch durchgesetzt wird. Handeln, welches durch seine korrekte Weise als solches bestimmt ist, ist die Weise menschlichen Verkehrs, dessen Grundform das Geben und Nehmen, genauer das Zusammenspiel von Fordern, Geben, Prüfen und Nehmen ist.“ (Löhrer 2003, 389)

<sup>427</sup> Vgl. Löhrer 2003, 401ff.

<sup>428</sup> Theoretische Gegenstücke zu diesen wären bloßes Raten und Fehlschlüsse.

berechtigterweise Zweifel daran besteht, ob Tiere über praktisches Wissen verfügen können. Ich denke, dass Korrektheit in dem Falle dieser Verhaltensweisen nur durch Beobachter festgestellt werden kann, die über ein solches praktisches Wissen verfügen. Für diese macht es Sinn, bei tierischem Verhalten korrekte von inkorrekten Verhaltensmustern zu unterscheiden, ebenso kanonische und nicht-kanonische Formen festzustellen und zu beschreiben. Auch hier lässt sich wohl an einem einzelnen Tier, in Isolation von allen anderen, nicht feststellen, ob sein Verhalten korrekt oder inkorrekt ist. Die Abstimmung der Mittel auf ihre Zwecke übernimmt im Falle von Tieren letztlich der evolutionäre Vor- oder Nachteil, der sich durch die jeweiligen Verhaltensweisen ergibt. Inkorrekte Verfahren sind evolutionäre Sackgassen. Auch wenn Tiere nicht über eine Repräsentation ihres praktischen Wissens verfügen, so ist dieses Wissen in ihrer evolutionären Abstimmung auf bestimmte Reiz-Reaktions-Schemen zum großen Teil fest eingeschrieben, andere Verhaltensweisen werden erst erlernt. Auch tierisches Verhalten kennt außerdem Methoden der Überprüfung der Korrektheit von Verhalten. Eine Unterwerfungsgeste eines Tieres in einem Rangkonflikt muss von dem anderen Tier durch entsprechendes Anschlussverhalten validiert werden. Tiere, deren Unterwerfungsgesten nicht korrekt funktionieren, werden dabei letztlich evolutionär selektiert. Natürlich ist diese Form der Überprüfung verschieden von der Prüfung von Handlungen durch menschliche Akteure, welche über geeignete Anschlusshandlungen nachdenken und sich bewusst auf gemeinschaftlich geteilte Normen in ihrer Überprüfung beziehen können. Hier möchte ich nur zeigen, dass es, bei Berücksichtigung geeigneter Einschränkungen, durchaus Sinn macht, auch bei Verhalten niederer Stufe von Korrektheit zu sprechen und praktisches Wissen als starr oder biologisch repräsentiert zu betrachten.<sup>429</sup> Die Lektion, die abweichende Kausalketten für die Diskussion über die Intentionalität von Handlungen bereithalten, besteht darin, dass die Intentionalität von Handlungen im Sinne der Korrektheit keine Eigenschaft ist, deren Vorliegen durch Referenz auf individuelle mentale Zustände von Akteuren wie Wünsche und Überzeugungen zu entscheiden ist. Es hängt nicht nur von den mentalen Zuständen von Akteuren ab, ob das, was sie tun, intentional ist oder nicht. Korrektheit ist ein Begriff, der keinen Sinn ergibt ohne den Verweis auf Maßstäbe, welche Vergleiche zulassen. Genauso wenig wie der Urmeter sich selbst messen kann, genauso wenig können individuelle mentale Eigenschaften von sich aus als Maßstab fungieren, um über ihre eigene Angemessenheit in Relation zu Handlungen zu befinden. Handlungstheorien, ob kausal oder teleologisch, welche diesen Umstand nicht in ihre Theorien integrieren, werden aufgrund der Möglichkeit abweichender Kausalketten keine befriedigende Theorie der Intentionalität von Handlungen im Sinne von Korrektheitsbedingungen formulieren können.

---

<sup>429</sup> Vgl. die Diskussion bei Spitzley 2009 über Willensschwäche bei Tieren.

## IV.5 Willensschwäche und Pläne

Donald Davidson hat in seinen ursprünglichen handlungstheoretischen Schriften ein reduktionistisches Programm für die drei Verwendungsweisen von ‚Intention‘ in Bezug zu Handlungen verfolgt. Handlungen sind dann intentional (absichtlich), wenn sie von primären Gründen verursacht sind. Eine Konstruktion eines primären Grundes liefert dabei zugleich die Intention, mit der eine Person handelt, und Pläne, etwas zu tun, sollten sich ebenso auf handlungsleitende Intentionen zurückführen lassen. Dieses Programm scheiterte aus seiner Sicht:

„When I wrote Essay 1 [*Action, reasons and causes*] [I] believed that of the three main uses of the concept of intention distinguished by Anscombe (acting with an intention, acting intentionally, and intending to act), the first was the most basic. Acting intentionally, I argued in Essay 1, was just acting with some intention. That left intending, which I somehow thought would be simple to understand in terms of the others. I was wrong. When I finally came to work on it, I found it the hardest of the three; contrary to my original view, it came to seem the basic notion on which the others depend; and what progress I made with it partially undermined an important theme in Essay 1 – that ‘the intention with which the action was done’ does not refer to an entity or state of any kind.” (Davidson 2001, xvii)

In seinen späteren Texten expliziert er ‚Intentionen‘ als evaluative Urteile besonderer Art, nämlich all-out oder nicht-konditionale Werturteile. Diese sind nicht hinreichend dafür, dass eine Person etwas intentional tut, da jemand Pläne fassen kann, die sie nie in die Tat umsetzt. Aber auch die Charakterisierung einer Handlung als ‚intentional‘, also absichtlich, bereitet Schwierigkeiten, da das Problem abweichender Kausalketten zeigt, dass nicht jede Handlung, die das beabsichtigte Handlungsergebnis hervorbringt, absichtlich ist. Die strikte Verbindung zwischen den drei Gebrauchsweisen von ‚Intention‘ war damit gebrochen.

Im Gefolge dieser Abschwächungen der ursprünglichen Theorie setzte auch ein gesteigertes Interesse an der Analyse von Plänen als eigenständige Entitäten ein. Vor allem Michael Bratman hat in *Intention, Plans, and Practical Reason* (1987) eine Interpretation von auf die Zukunft gerichteten Intentionen als Pläne gegeben. Akteure verfolgen nicht einzelne, voneinander isolierte Absichten, sondern formen Pläne, also Handlungssequenzen und Handlungsabfolgen, welche in logischen und zeitlichen Ordnungen zueinander stehen. Neben der Rationalisierungsbeziehung, in der einzelne Handlungen zu ihren Gründen stehen, können auch Rationalitätsbedingungen für Pläne formuliert werden.

„The central fact is that we are planning agents. We frequently settle in advance on more or less complex plans concerning the future, and these plans guide our later conduct. [...] As planning agents, we have two central capacities. We have the capacity to act purposively; and we have the capacity to form and execute plans.” (Bratman 1987, 2)

Pläne helfen rationale Wesen auf zwei Weisen: einerseits sind sie Speichermedien für praktische Überlegungen, welche daher zu Zeitpunkten durchgeführt werden können, in denen die dafür nötige Ruhe und Zeit vorhanden ist. Andererseits erlauben sie die Koordination und Integration

verschiedener komplexer Handlungsabläufe.<sup>430</sup> Charakteristisch ist auch, dass Pläne, die zukünftiges Handeln betreffen, auf Typen von Handlungen gerichtet sind und nicht auf partikuläre Handlungen. Dies deshalb, da zum Zeitpunkt des Fassens eines solchen Planes noch keine partikuläre Handlung existiert, auf die sich der Plan richten könnte.

In welchem Bezug stehen nun Pläne zur Willensschwäche? Normalerweise haben willensschwache Personen nicht den Plan, willensschwach zu handeln, sondern sie handeln und zeigen dabei Willensschwäche. Könnte jedoch eine Person überhaupt einen Plan fassen, diesen umsetzen, und dabei willensschwach handeln? Folgendes Argument spricht gegen diese Möglichkeit:

- (P1) Wenn eine Person einen Plan hat und diesen erfolgreich umsetzt, dann handelt sie so, wie sie es geplant hat.
- (P2) Wenn eine Person willensschwach handelt, dann handelt sie nicht so, wie sie es geplant hat.
- (K) Also: Eine Person kann nicht willensschwach handeln, dabei einen Plan haben und diesen erfolgreich umsetzen.

Dieses Argument kann zumindest aus zwei Gründen kritisiert werden: 1) Prämisse 2 beschreibt nur Fälle von exekutiver, nicht von motivationaler Willensschwäche. 2) Das Argument beruht auf einer unrechtmäßigen Äquivokation: der Plan in Prämisse 1 ist nicht notwendigerweise derselbe wie in Prämisse 2.

1) Motivationale Willensschwäche ist der Fall, bei dem ein Akteur wider sein eigenes bestes oder besseres Urteil handelt, also nicht die Handlung beabsichtigt, die er selbst meint, dass er beabsichtigen sollte. Dies ist zu unterscheiden von exekutiver Willensschwäche, in der jemand nicht so oder nicht in der Weise handelt, wie er es geplant hat. Da Selbstreferenz auch in anderen Kontexten für Verwirrung und Paradoxien sorgt, ist es nicht überraschend, dass dies auch bei Willensschwäche der Fall ist. Folgende Fälle sind deshalb gemäß der Unterscheidung zwischen motivationaler und exekutiver Willensschwäche zu unterscheiden: a) Kann eine Person urteilen, dass es, alles in allem am besten wäre, motivational willensschwach zu handeln, und dies auch tatsächlich tun? b) Kann eine Person den Plan fassen, willensschwach zu handeln, und dies auch erfolgreich tun?

In Bezug zu a) lässt sich feststellen, dass hier tatsächlich ein Urteil vorliegt, dass ein Handlungsziel vorgibt, das unmöglich zu realisieren ist. Denn: Angenommen, die Person urteilt, dass es am besten wäre, motivational willensschwach zu handeln, d.h. nicht das zu intendieren und tun, was sie selbst als bestes ansieht. Wie sollte eine Person eine solche paradoxe

---

<sup>430</sup> Vgl. Bratman 1987, 2ff.

Aufforderung erfüllen? Handelt sie so, wie sie urteilt, also willensschwach, dann tut sie etwas, das diesem Urteil widerspricht, also handelt sie nicht willensschwach. Handelt sie nicht so, wie es das Urteil vorsieht, dann handelt sie willensschwach, also genau so, wie es das Urteil vorsieht, was der Annahme widerspricht. Also: die Person kann nicht urteilen, dass es am besten wäre, motivational willensschwach zu handeln und dies erfolgreich umsetzen.<sup>431</sup>

In Bezug auf b) besteht die Frage darin, ob es vorstellbar ist, dass eine Person den Plan fassen kann, etwas zu tun, obwohl sie urteilt, dass sie nicht in dieser Weise handeln sollte. Dies ist dann nicht vorstellbar, wenn ein Plan, etwas zu tun, notwendigerweise mit dem besten Urteil übereinstimmt oder identisch mit diesem Urteil ist. Denn dann wäre jeder Fall von 2) auch ein Fall von 1), der jedoch, wie gezeigt, nicht zu realisieren ist. Ich habe bereits mehrere Theorien diskutiert, welche Intentionen als Pläne eng an beste Urteile knüpfen.<sup>432</sup> Hier stellt sich die Frage, ob – gewissermaßen aus der anderen Richtung – im Begriff des Planes etwas enthalten ist, dass Pläne notwendig mit besten Urteilen übereinstimmen lässt.

Ein Plan zu haben umfasst normalerweise eine propositionale Komponente und eine Akt-Komponente, welche die Haltung zum propositionalen Inhalt des Planes beschreibt. Einen Plan zu haben wie ‚Ich plane F‘ beinhaltet zumindest den Glauben zu haben, dass eine Chance besteht, F zu tun. Widerspricht ein solcher Glaube einem besseren Urteil mit dem Inhalt, F nicht zu tun? Nur dann, wenn die Person zusätzlich glaubt, dass ein solches Urteil notwendig mit dem Plan verbunden ist, F nicht zu tun, der dem Plan, F zu tun, also direkt widersprechen würde. Doch eine Person muss nicht über diesen Glauben verfügen.<sup>433</sup> Beinhaltet einen Plan zu fassen außerdem zu glauben, einen Grund zu haben, entsprechend zu handeln? Oder anders: Kann eine Person einen Plan haben, etwas zu tun, ohne zu glauben, dass sie einen Grund hat für das, was sie zu tun plant? Die Person fällt schließlich ein Urteil derart, dass die besseren Gründe gegen die willensschwache Handlung sprechen. Wie kann sie dann gleichzeitig einen Grund für die Ausführung der willensschwachen Handlung sehen? Klar ist, dass in dem Begriff des Planes ohne weitere Annahme keine notwendige Verbindung zu Handlungsgründen besteht. Die Konsequenz davon ist, dass eine Person demnach einen Plan haben könnte, von dem sie selbst nicht überzeugt ist, dass dieser gerechtfertigt ist. Ich halte fest, dass ohne weitere Annahmen nichts

---

<sup>431</sup> Vgl. Arpaly 2000, 490. „Similarly, "I, the agent, should act against my best judgment" – itself a best judgment rather than advice – is flagrantly absurd. Acting against one's best judgment is not something that one can, as a result of deliberation, resolve to do; one just changes one's best judgment as to what to do." (Arpaly 2000, 490)

<sup>432</sup> Vgl. Kap. III.

<sup>433</sup> Ein spezielles Argument besagt, dass es moralisch besser wäre, die Existenz von Willensschwäche zu verneinen, da so eine zusätzliche Möglichkeit zur Ausrede, weshalb jemand nicht das tut, was er meint tun zu sollen, wegfallen. Ob Personen tatsächlich moralischer handelten, wenn sie davon ausgingen, dass es Willensschwäche nicht gibt, ist eine empirische Behauptung. Abgesehen davon sprechen gute Gründe dagegen, Theorien aus scheinbar moralischen Gründen abzulehnen, da dies dem formalen Ziel des Glaubens nach Wahrheit widerspricht.

dagegen spricht, dass Personen Pläne fassen können, die nicht mit ihren evaluativen Urteilen übereinstimmen.

2) Der zweite Einwand gegen das obige Argument betrifft die mögliche Äquivokation von ‚Plan‘ in den beiden Prämissen: Wenn die Prämissen auf jeweils verschiedene Pläne des Akteurs referieren, dann folgt die Konklusion nicht. Exekutive Willensschwäche besteht darin, wider eigene Pläne intentional zu handeln. Der ursprüngliche Plan wird also nicht umgesetzt. Dies heißt aber nicht, dass die Person nicht nach einem *anderen* Plan handeln könnte. Ein Widerspruch würde sich nur dann einstellen, wenn willensschwach zu handeln impliziere, keinen Plan erfolgreich umzusetzen. Das Argument würde dann lauten:

- (P1) Wenn jemand einen Plan hat und diesen erfolgreich umsetzt, dann handelt sie so, wie sie es geplant hat.
- (P2) Wenn jemand willensschwach handelt, dann handelt sie in keiner Weise so, wie sie es geplant hat.
- (K) Also: Eine Person kann nicht willensschwach handeln, dabei einen Plan haben und diesen erfolgreich umsetzen.

In dieser Form folgt die Konklusion aus den Prämissen, jedoch ist die Charakterisierung von Willensschwäche in Prämisse 2 problematisch. Die Person, die exekutive Willensschwäche zeigt, handelt gegen einen Plan, aber dies heißt nicht, dass sie dabei nicht anhand eines anderen Planes handeln könnte. Folgende positive Gründe sprechen dafür, willensschwachen Personen die Fähigkeit zuzuschreiben, Pläne für ihre willensschwachen Handlungen zu fassen und nach ihnen zu handeln: *Erstens* gibt es willensschwache Handlungen, welche komplizierte Vorbereitungs-handlungen benötigen und deren Ausführung ein hohes Maß an Planung voraussetzen. Ein Beispiel: Raimund könnte glauben, dass es am besten wäre, in einer Woche nach New York zu fliegen, doch fasst er stattdessen den Plan, nach Buenos Aires zu fliegen. Er fasst diesen Plan am Montag, kauft seine Tickets am Mittwoch, und packt seine Koffer am Sonntag. Er sagt außerdem mit diesem Plan unvereinbare Termine ab und sieht zu, dass sein Plan, nach Buenos Aires zu fliegen, mit seinen anderen Plänen zusammenstimmt. *Zweitens* können Personen willensschwache Pläne fassen, die sie nie in die Tat umsetzen, weil sich z.B. wesentliche Umstände ihres Handelns ändern, oder weil sie ihre Willensschwäche überwinden. Raimund könnte am Montag den Plan fassen, nach Buenos Aires zu fliegen, jedoch am Mittwoch, wenn er die Tickets kaufen wollte, diesen Plan ändern und nun doch welche nach New York lösen.

Zusammenfassend stelle ich fest, dass keine guten Gründe bestehen, willensschwachen Personen die Ausbildung und Umsetzung von Plänen abzusprechen. Einige willensschwache Handlungen können auch nur geplant stattfinden, da sie Vorbereitungshandlungen benötigen und

komplizierte Handlungsabläufe umfassen können.<sup>434</sup> Zudem könnten die Pläne für willensschwache Handlungen auch wieder geändert oder nie in die Tat umgesetzt werden. Der Widerstand, willensschwache Handlungen als teilweise geplant aufzufassen, rührt daher, dass normalerweise Pläne mit den Urteilen darüber, was am besten zu tun ist, übereinstimmen. Da Pläne jedoch auch spontan oder durch Gewohnheit gebildet werden können, spricht nichts dagegen, dass auch willensschwache Personen nach Plänen handeln können. Die Ausarbeitung des Planes erfordert praktische Überlegung, und die willensschwache Handlung entspricht in diesen Fällen in dieser Weise der praktischen Überlegung. Doch das Ziel des Planes, der Zweck, auf den die verschiedenen Teilhandlungen organisiert sind, ist nicht notwendigerweise ein rationales Produkt praktischer Überlegung, und entspricht in diesem Sinne nicht der praktischen Überlegung. Eine willensschwache Handlung kann daher in einem Sinn als überlegt bezeichnet werden, im Sinne von planvoll, und gleichzeitig als unüberlegt, im Sinne der Nichtübereinstimmung mit den evaluativen Urteilen der Person.

## IV.6 Exekutive Willensschwäche

In *Willing, Wanting, Waiting* (2009) entwirft Richard Holton eine Theorie von exekutiver Willensschwäche, welche auf den ‚Willen‘ als eigenständige und von Glaubenshaltungen und Wünschen relativ unabhängige Instanz aufbaut.<sup>435</sup> Dabei macht Holton Gebrauch von empirischen Befunden aus der Psychologie, welche mit seinem Konzept gut übereinstimmen. Personen, so seine These, verfügen über ein bestimmtes Maß an Willenskraft, eine dynamische Größe, welche entscheidend ist für die Umsetzung und Realisierung von Plänen und Vorsätzen.<sup>436</sup> Willensschwäche besteht darin, dass eine Person gegen einen Vorsatz (resolution) handelt, dessen Funktion genau darin besteht, Intentionen vor kurzfristige und durch Begierden hervorgerufene Änderungen zu schützen.

---

<sup>434</sup> Bereits Austin hat darauf hingewiesen, das einer Versuchung nachgeben nicht gleichzusetzen ist mit dem völligen Verlust von Kontrolle: „I am very partial to ice cream, and a bombe is served divided into segments corresponding one to one with the persons at High Table: I am tempted to help myself to two segments and do so, thus succumbing to temptation and even conceivably (but why necessarily?) going against my principles. But do I lose control of myself? Do I raven, do I snatch the morsels from the dish and wolf them down, impervious to the consternation of my colleagues? Not a bit of it. We often succumb to temptation with calm and even with finesse.” (Austin 1956–57, 24)

<sup>435</sup> Richard Holton unterscheidet zwischen ‚Akrasia‘ und ‚Willensschwäche‘, was ich mit motivationaler und exekutiver Willensschwäche bezeichnet habe. Vgl. Kap. I.3.

<sup>436</sup> „The central idea is that of an executive capacity. It is a capacity needed by cognitively limited creatures like us, acting in a complex world that is full of uncertainties and temptations. Often we have to act in the absence of a judgement of what is best. That is why we need a capacity to make choices that is independent of our judgements. Even if we can make judgements about what is best, they are often corrupted by ensuing temptation. Temptation does not just change what we want; it also changes our judgements. So again we need a capacity to stick by our intentions that is independent of judgement. The capacities that provide such abilities form the core of our talk of strength and weakness of the will, and of will power.” (Holton 2009, x)

„[...] [W]eakness of will is unreasonable revision of a contrary inclination defeating intention (a resolution) in response to the pressure of those very inclinations.“ (Holton 2009, 78)

Eine entscheidende Frage für diesen Ansatz ist, wie unvernünftige von vernünftigen Änderungen der Vorsätze unterschieden werden können, da nicht jede Änderung eines Vorsatzes als willensschwach bezeichnet wird.<sup>437</sup> Schließlich gibt es oftmals gute Gründe, einen zuvor gefassten Vorsatz zu ändern, falls neue Informationen dies nahelegen. So ist ein Bergsteiger nicht willensschwach, wenn er den Vorsatz ändert, als erster einen bestimmten Berg zu besteigen, wenn er erfährt, dass er schon längst von jemand anderen bestiegen wurde. Holton entwirft keine formale und vollständige Theorie, anhand dessen die Frage der Vernünftigkeit von Änderungen eigener Vorsätze entschieden werden könnte.<sup>438</sup> Generell verteidigt er den Ansatz, dass die Rationalität, Vorsätze einzuhalten, darin besteht, dass einen Vorsatz zu fassen einer Person einen Grund gibt, keine weiteren Überlegungen auch bei einsetzenden starken Begierden über das eigene Handeln anzustellen.<sup>439</sup> Resolute Akteure betrachten ihre Überlegungen als abgeschlossen und sind daher weniger als irresolute bzw. willensschwache Akteure geneigt, ihre Vorsätze zu ändern. Das Fassen eines Vorsatzes gibt dabei keinen Grund, entsprechend dem Vorsatz zu handeln, sondern nur einen Grund, keine weiteren Überlegungen über das eigene Handeln anzustellen.<sup>440</sup> Die Tendenz bzw. Disposition, den Vorsatz vor neuen Überlegungen zu schützen, ist dabei selbst nicht reflexiv oder deliberativ, ansonsten wäre die Person ständig mit der Überlegung beschäftigt, ob sie nun neu überlegen sollte oder nicht, was wiederum dem Zweck einer Entlastung von epistemischen Ressourcen durch das Fassen eines Vorsatzes widersprechen würde. Willensstärke in diesem Sinne hilft demnach nicht nur, eigene Vorsätze auch tatsächlich umzusetzen, sondern schützt dabei auch vor diachronen Formen motivationaler Willensschwäche, da die Person nicht in die Lage gerät, ihr besseres Urteil zu überdenken und es aufgrund gerade auftretender Wünsche und Begierden zu ändern. Sturheit wiederum markiert die Grenze, wo es für Personen rationaler wäre, einmal gefasste Intentionen aufgrund neuer Umstände zu überdenken, anstatt sie einfach in die Tat umzusetzen.<sup>441</sup>

---

<sup>437</sup> Vgl. aber die Kritik von Dylan Dodd, der dafür argumentiert, nicht nur unvernünftiges, sondern jedes Handeln wider eine bestehende Intention als willensschwach zu verstehen. „In contrast to Holton, I claim that an agent is weak-willed *whenever* (s)he act contrary to one of her intentions, not just when she should have stuck with the relevant intention.“ (Dodd 2009, 45)

<sup>438</sup> Für entsprechende Daumenregeln vgl. Holton 2009, 75.

<sup>439</sup> Vgl. Holton 2009, ch. 7.

<sup>440</sup> Zum ansonsten drohenden Problem des ‚bootstrapping‘ von Gründen auf Basis von Intentionen vgl. die Diskussion bei Bratman 1987; Holton 2009, 143–148; Broome 2001.

<sup>441</sup> Vgl. Holton 2009, 79f.

## Zusammenfassung

Die Frage nach der Intentionalität von willensschwachen Handlungen kann auf drei Weisen verstanden werden: als Frage nach deren handlungsleitenden Intention, nach deren Absichtlichkeit und nach deren Planbarkeit. Gegen die Existenz von handlungsleitenden Intentionen bei willensschwachen Handlungen sprechen drei Prinzipien: eine Rückwärts-Verbindung (backward connection) zwischen intentionalen Handlungen und besseren Urteilen, eine Vorwärts-Verbindung (forward connection) zwischen besseren Urteilen und Intentionen, und Gründe-Theorien intentionalen Handelns. Für die Existenz einer backward connection sprechen entscheidungstheoretische Grundannahmen zur Interpretation von Handlungen. Doch sind diese nur für Akteure gültig, deren Präferenzstruktur bereits rational ist und einigen grundlegenden Axiomen genügt. Außerdem fußt diese Theorie auf einer behavioristischen Interpretation von Präferenzen, welche das Verhältnis des Akteurs zu seinen eigenen Wünschen in Form von evaluativen Urteilen methodisch ausblendet. Dagegen könnte eine Unterscheidung zwischen starken und schwachen Wertungen den Unterschied zwischen auf moralischen Überlegungen und Klugheitsüberlegungen fußenden Evaluationen und solchen, die auf bloßen Wünschen basieren, markieren. Für die forward connection sprechen die einleuchtenden Prinzipien, das Personen das tun, wozu sie am meisten motiviert sind, und am meisten motiviert sind zu tun, was sie am besten bewerten. Gegen letzteres spricht vor allem, dass hier der Zusammenhang zwischen Evaluation und Motivation zu eng gefasst ist, und dass Intentionen, etwas zu tun, verschieden sind von evaluativen Urteilen über Handlungsoptionen. Für die Gründe-Theorie spricht, dass jede Handlung aus zumindest einem Grund vollzogen wird. Dies schließt jedoch nicht aus, dass Personen Handlungen aus Gründen vollziehen, die sie selbst als schlecht bewerten. Daher braucht es eine Theorie praktischer Gründe, welche zeigt, wie solche Gründe bestehen bleiben, auch wenn sie aus Sicht der Person relativ zu anderen Gründen die Handlung nicht rechtfertigen. Praktische Gründe können auf verschiedene Weisen verstanden werden. Hier schlage ich einen Mittelweg zwischen wunsch- und wertebasierten Theorien vor: einige praktische Gründe basieren auf Werten, andere auf Wünschen von Personen. Viele Handlungen machen nur Sinn, wenn die Person spezielle Wünsche besitzt, welche durch die Handlung erfüllt werden: Niemand hat einen Grund, Schach zu spielen, wenn kein besonderes Interesse an Schach besteht. Dagegen behaupten moralische Gründe eine relative Unabhängigkeit gegenüber der Existenz solcher Wünsche. Für die Absichtlichkeit von Handlungen im Sinne eines Korrektheitskriteriums lassen sich drei Stufen zielgerichteten Verhaltens unterscheiden: basale Intentionalität, präreflexive Intentionalität und reflexive Intentionalität. Willensschwache Handlungen befinden sich auf der Stufe zwischen präreflexiver und reflexiver Intentionalität. Trotz der Reflexion auf eigene Handlungsgründe handelt die Person so, als hätte sie gar nicht

überlegt, mithin in einem präreflexiven Modus. Die Frage der Planbarkeit von willensschwachen Handlungen ist differenziert zu beantworten. Willensschwache Handlungen können geplant verlaufen, wenn deren Komplexität die Koordination verschiedener Teilhandlungen verlangen. Dagegen können Personen nicht planen, motivational willensschwach zu handeln: ein solches besseres Urteil widerspräche sich selbst, wodurch das entsprechende Handeln nicht mehr als Handeln wider das alles in allem bessere Urteil verstanden werden könnte. Das Phänomen exekutiver Willensschwäche verweist auf die Existenz einer nondeliberativen Fähigkeit rationaler Akteure, ihre Intentionen in Handlungen umzusetzen und ihren Vorsätzen und Plänen für die Zukunft Stabilität zu verleihen.

## V Problemfeld Verantwortung

In diesem Kapitel diskutiere ich die Frage, ob und weshalb es gerechtfertigt ist, Personen Verantwortung für ihre Willensschwäche und willensschwachen Handlungen zuzuschreiben. Zuerst analysiere ich den Begriff der Verantwortung als vierstellige Relation, deren Funktion in Zuschreibung und Zuweisung von Normen zu Personen besteht (1). Danach untersuche ich verschiedene Kriterien für die Zuschreibung von Verantwortung für Handlungen (2). Es folgt die Frage nach dem Verhältnis von Zwang und Willensschwäche (3). Das Ergebnis ist, dass Verantwortung für willensschwache Handlungen zu Recht zugeschrieben wird, da vergleichbare Personen mit einem für die jeweilige Gesellschaft als Standard vorausgesetzten Grad an Selbstkontrolle die Handlungen hätten vermeiden können.

### V.1 Begriff der Verantwortung

Der Begriff der Verantwortung wird auf verschiedene Weisen verwendet.<sup>442</sup> Hier interessiere ich mich ausschließlich für seine Verwendung als Zuschreibungsbegriff.<sup>443</sup>

„Dieser Sinn und Zweck [des Verantwortungsbegriffs] besteht darin, eine Beziehung zwischen (moralischen) Subjekten und (moralischen) Objekten herzustellen. Dies ist immer dann gelungen, wenn das Objekt einem Subjekt zugerechnet wird: gleichgültig, ob es sich bei diesem Objekt um eine Handlungsfolge (retrospektive Zurechnung) oder um einen zu erhaltenden oder herzustellenden Zustand handelt (prospektive Zurechnung). Aufgabe einer Theorie der Verantwortung ist es dementsprechend, Kriterien für solche Zurechnungen zu entwickeln und zu begründen.“ (Bayertz 1995, 64f.)

Peter Cane unterscheidet zudem zwischen vier Funktionen, welche unsere Praktiken der Verantwortungszuschreibung erfüllen:<sup>444</sup> 1) Ontologische Funktion: die Zuschreibung von Handlungen und Handlungsergebnissen zu Akteuren (Individuen oder Gruppen); 2) Explanatorische Funktion: die Erklärung von Sachverhalten, vor allem hinsichtlich ihrer kausalen Genese; 3) Normative Funktion: die Präzisierung normativ richtigen Verhaltens, vor allem mit Blick auf das zukünftige Verhalten; 4) Evaluative Funktion: die Evaluation vergangenen Verhaltens. Verantwortung ist demnach ein komplexer Begriff, der verschiedene Funktionen innerhalb des gesellschaftlichen Zusammenlebens erfüllt und daher verschiedene Aspekte und Dimensionen aufweist.

„Verantwortung“ verstehe ich als einen Zuschreibungsbegriff, dessen Funktion es ist, eine Norm einem Subjekt hinsichtlich eines Sachverhaltes konkret zuzuweisen. Der Begriff der Verant-

---

<sup>442</sup> Für eine Übersicht über die philosophische Diskussion verschiedener Konzeptionen und Formen von Verantwortung vgl. Fischer 1999; Hart 2008; Cane 2002. Letzterer unterscheidet einerseits zwischen prospektiver und retrospektiver Verantwortungszuschreibung. Außerdem zwischen kausaler Verantwortung, Fähigkeitsverantwortung (capacity responsibility) und Rollenverantwortung.

<sup>443</sup> Im Unterschied zur Verwendung als Wertbegriff. Als solcher bezeichnet er die (besondere) Erfüllung von Verantwortungszuschreibungen, z.B. in ‚X hat verantwortlich gehandelt‘ oder ‚X ist ein verantwortungsvoller Mensch‘. Vgl. Cane 2002, 34.

<sup>444</sup> Vgl. Cane 2002, 57.

wortung lässt sich dabei als mehrstellige Relation zwischen einem Verantwortungssubjekt, einem Gegenstand, einer Instanz und einem Normhintergrund analysieren:<sup>445</sup>

Jemand (Subjekt) ist für etwas (Gegenstand) vor oder gegenüber jemandem (Instanz) aufgrund bestimmter normativer Standards (Normhintergrund) verantwortlich.

Für jede der Relationsglieder lassen sich umfangreiche philosophische Debatten finden: Was sind mögliche Subjekte von Verantwortung, was mögliche Objekte, welchen Instanzen gegenüber verantworten sich Subjekte und aufgrund welcher normativer Standards wird Verantwortung zugeschrieben? Einfach lässt sich dieses Schema anhand legaler Verantwortung verstehen: ein Rechtssubjekt, z.B. eine natürliche Person, ist für ihre Handlung, die sich unter einen Tatbestand subsumieren lässt, z.B. einen Diebstahl, gegenüber der Gesellschaft in Gestalt des Richters verantwortlich aufgrund entsprechender Rechtsnormen, niedergelegt im Strafrecht. Hinzu kommen noch Diskussionen um metaphysische Voraussetzungen von Verantwortung, wie z.B. Diskussionen um Determinismus und Indeterminismus und das sogenannte Vereinbarkeitsproblem.<sup>446</sup> Hier diskutiere ich diese metaphysischen Fragen nicht und frage nur, weshalb Personen zu Recht Verantwortung für ihre Willensschwäche und willensschwachen Handlungen zugeschrieben wird, *falls* es Verantwortung im Sinne einer Konzeption von Verantwortung gibt, wie sie die alltägliche Erfahrung nahelegt. Da ich Willensschwäche auf der Ebene des Individuums, und nicht von Kollektiven untersuche, kommt als Verantwortungssubjekt nur das jeweilige Individuum in den Blick. Ob das gesellschaftliche Umfeld teilweise für die Ausbildung von Willensschwäche einer Person verantwortlich ist, lasse ich hier offen, obwohl dies in gewissem Umfang sicherlich zutrifft.<sup>447</sup> Die Objekte der Verantwortung im Zusammenhang mit Willensschwäche sind willensschwache Handlungen und Willensschwäche als Charakterzug. Ob hier eine Verantwortungszuschreibung reduzierbar ist auf die andere ist fraglich. Als Instanzen kommen vor allem Personen vor, die in einem engen Verhältnis zur willensschwachen Person stehen, und die Person selbst im Sinne einer Idee von Selbstverantwortung. Die normativen Standards sind verschieden: genauer setze ich mich mit diesen in Kap. VI auseinander. Sie umfassen moralische, prudentielle und möglicherweise auch ästhetische Normen. Die Normen, die eine willensschwache Person durch ihr Handeln verletzt (oder erfüllt), können vielfältig sein, abhängig vom jeweiligen Handlungstyp, der mit der Handlung realisiert wird und den tatsächlichen kausalen Folgen des Handelns. Normativ betrachtet dreht sich Willensschwäche

---

<sup>445</sup> Vgl. Werner 2006.

<sup>446</sup> Die Vereinbarkeit moralischer Verantwortung mit der Vorstellung eines deterministischen Weltverlaufes, vgl. Keil 2007. Oftmals mangelt es diesen Diskussionen an einem hinreichend klaren Verständnis von Determinismus und Kausalität.

<sup>447</sup> Vgl. Rorty 1997.

jedoch im Kern um den Wert von Selbstrespekt und fundamentale Normen praktischer Rationalität.

In der Philosophie wird oftmals zwischen moralischer Verantwortung und anderen Formen von Verantwortung unterschieden, wobei angenommen wird, dass diese strikt voneinander getrennt wären. Moralische Verantwortung impliziere Schuldhaftigkeit, und Schuldhaftigkeit impliziere intentionales Handeln bzw. eine bestimmte Qualität des Willens, das dem Handeln zugrunde liege.<sup>448</sup> Diese Theorien konzentrieren sich zudem auf Eigenschaften des Handelnden, welche als Kriterien für die Zuschreibung von Verantwortung dienen können. Diese Kriterien sind sicherlich wichtig, doch geht die Fixierung auf die Rolle des Akteurs auf Kosten eines besseren Verständnisses des relationalen Charakters von Verantwortung und der verschiedenen Funktionen, welche diese Zuschreibung nicht nur im rechtlichen, sondern auch moralischen Bereich, erfüllt.

„Responsibility is not just a function of the quality of will manifested in conduct, nor of the quality of that conduct. It is also concerned with the interest we all share in security of person and property, and with the way resources and risks are distributed in society. Responsibility is a relational phenomenon.” (Cane 2002, 109)

In diesem Sinne verstehe ich auch die Verantwortung für Willensschwäche und willensschwache Handlungen als im Kern relational und nicht mit starken theoretischen Vorannahmen belastet.

## **V.2 Kriterien für die Zuschreibung von Verantwortung**

In Kap. IV habe ich die Frage nach der Intentionalität willensschwacher Handlungen auf verschiedene Weise untersucht. Dabei habe ich festgestellt, dass willensschwache Handlungen Fälle von nicht geglückter reflexiver Selbstbestimmung sind, da aus Sicht der Person die Gründe, die sie für ihre willensschwache Handlung hat, ihre Handlung im Vergleich zu ihren umfassenden Handlungsgründen nicht rechtfertigen. Dennoch sind willensschwache Handlungen intentional: sie sind kontrolliert, zielorientiert, und durch intentionale Einstellungen des Akteurs partiell begründet. Zudem können Personen Pläne für ihre willensschwachen Handlungen fassen und ihnen gemäß handeln. Für eine Position, die ich hier als klassisch bezeichne, ergibt sich aus der Intentionalität der entsprechenden Handlungen sogleich ein Grund für die Zuschreibung von Verantwortung: intentionale Handlungen sind nach diesem Modell die primären Objekte, für die Personen verantwortlich sind. Die Intentionalität der Handlung wird dabei als notwendige und hinreichende Bedingung für die Verantwortungszuschreibung gedacht: Eine Person ist für die Handlung H dann und nur dann verantwortlich, wenn sie H intentional (absichtlich) tut. Diese Verantwortung überträgt sich gemäß diesem Konzept transitiv auf die Folgen von intentionalen

---

<sup>448</sup> Vgl. die Kritik von Cane 2002, ch. 3.6, über ‚choice theories‘ und deren Akteurszentriertheit.

Handlungen. Im Unterschied dazu meine ich, dass die Intentionalität von Handlungen zwar sicherlich zentral für unsere Praxis der Verantwortungszuschreibung, doch weder a) hinreichend noch b) notwendig für diese ist.

a) Personen können etwas absichtlich tun, und dieses auch erfolgreich durchführen, und dennoch nicht für die Handlung und deren Folgen verantwortlich sein. Dies ist dann der Fall, wenn die Person zu der Handlung gezwungen wird. Ein Beispiel: eine Bankräuberin droht einem Bankangestellten, ihn zu erschießen, wenn er ihr nicht das geforderte Geld aushändigt. Der Bankangestellte übergibt das Geld. Die Geldübergabe ist intentional: sie ist zielgerichtet, steht unter der Kontrolle der Person, und folgt Gründen, welche die Person für die Handlung hat. Dennoch wird der Bankangestellte von seiner Verantwortung für die Geldübergabe entschuldigt und ist für sie nicht verantwortlich. Die Intentionalität der Handlung ist nicht hinreichend für die Verantwortungszuschreibung, da sie das Fehlen von Zwang voraussetzt.

b) Personen können auch für Tätigkeiten und Folgen von diesen verantwortlich sein, die sie nicht absichtlich herbeiführen. So ist jeder (genuine) Fehler, der einer Person unterläuft, nicht absichtlich. Niemand kann sich absichtlich verrechnen. Dennoch können Personen verantwortlich sein für Folgen, die aus Fehlern und fehlerhaften Handlungen entstehen. Dieser Umstand wird oftmals so eingeholt, dass in diesen Fällen die Person dann für ihre Handlungsfolgen verantwortlich ist, wenn es ihr möglich gewesen wäre, den Fehler zu vermeiden. Damit wechselt der Fokus von Kriterien, die an das mögliche Objekt der Verantwortung angelegt werden, zu solchen, denen das Verantwortungssubjekt genügen muss. Denn die Feststellung, dass eine Person fähig gewesen wäre, den Fehler nicht zu tun, richtet sich nach den Fähigkeiten des Akteurs, und nur teilweise nach den Eigenschaften der tatsächlich ausgeführten Handlung.

Bedeutsam ist in diesem Zusammenhang, dass die Zuschreibung von Verantwortung in diesen Fällen oftmals der fehlerhaften Handlung *vorausgeht*. Eine Person übernimmt die Verantwortung für die *Durchführung einer bestimmten Tätigkeit*. Tritt ein Fehler innerhalb dieses Verantwortungsbereiches auf, so liegt es an der Person, die Verantwortungszuschreibung für diesen Fehler zurückzuweisen, falls die Person meint, dass tatsächlich eine Ausnahme vorliege. Kurt Bayertz stellt dazu fest, dass soziologisch gesehen diese Verschiebung hin zur vermehrten prospektiven Zuschreibungen von Verantwortung ein Ergebnis der Industrialisierung ist, vor allem hinsichtlich der mit ihr verbundenen massenhaften Arbeitsteilung und Technisierung der Arbeitswelt.<sup>449</sup> Mit der prinzipiellen Zunahme von Risiken, die mit technischen Arbeitsprozessen verbunden sind,

---

<sup>449</sup> „Die extensive Verwendung des Verantwortungsbegriffs kann somit als ein Ausdruck der diffiziler gewordenen Steuerungsprobleme in modernen, funktional ausdifferenzierten Gesellschaften verstanden werden.“ (Bayertz 1995, 34)

und dem unklaren kausalen Anteil individueller Akteure am jeweiligen Endresultat, ergibt sich der Bedarf nach einer Regelung der Verantwortungszuschreibung, die das klassische Modell der retrospektiven Zuschreibung von einzelnen Handlungsfolgen ergänzt bzw. überlagert.

„Mit dem Verantwortungsbegriff in seiner nichtklassischen Bedeutung werden demgegenüber bestimmte positiv ausgezeichnete (erwünschte) Zustände umschrieben, ohne daß im einzelnen festgelegt werden muß, wie der jeweilige Verantwortliche diese Zustände herbeiführt oder aufrechterhält. Wenn jemand für eine Maschine „verantwortlich“ ist, dann hat er sicherzustellen, daß alle erforderlichen Maßnahmen ergriffen werden, um Schäden von ihr fernzuhalten und dennoch auftretende Schäden zu beseitigen: ohne daß die Schäden oder die erforderlichen Maßnahmen vorher spezifiziert werden müssen. [...] Spezifisch für „Verantwortung“ ist gerade, daß die Anforderungen vorher *nicht* formuliert werden können und daß daher im einzelnen offenbleiben muß, *wie* Verantwortung wahrgenommen wird.“ (Bayertz 1995, 33f.)

In einem klassischen Verständnis von Verantwortung spielt neben der Intentionalität auch die Verursachung eine zentrale Rolle: sie funktioniert als notwendige Bedingung für die Verantwortungszuschreibung. Nur die Folgen, die kausal aus dem Handeln einer Person folgen, sind auch der Person zuzuschreiben. Als hinreichend wird diese Verbindung meist nicht angesehen, da es leicht der Fall sein kann, dass eine Person kausal in ein Geschehen involviert ist, ohne Verantwortung für dieses zu tragen. Folgendes spricht gegen eine kausale Verknüpfung als notwendige Bedingung der Verantwortungszuschreibung:<sup>450</sup>

a) Unterlassungen: nach einem üblichen Verständnis kausaler Beziehungen sind Unterlassungen keine kausalen Ursachen. Wie kann eine Handlung, die nicht ausgeführt wird, etwas verursachen? Nach einer ereignisontologischen Auffassung von Handlungen könnte man einwenden, dass falls Unterlassungen Handlungen sind, sie auch irgendwelchen Ereignissen entsprechen müssen. Diese könnten dann als Ursache angesehen werden. Interessant ist hierbei, dass auch diesen Fällen die Verantwortungszuschreibung der Handlung vorausgeht: wenn nicht klar wäre, welche Handlung in dieser Situation von der Person gefordert ist, für deren Ausführung sie demnach verantwortlich wäre, könnte ihr tatsächliches Handeln nicht dieser Erwartung zuwiderlaufen bzw. könnte sie nicht die geforderte Handlung unterlassen. Ebenso sind Fälle möglich, in denen die Person die entsprechenden Handlungen nicht intentional (absichtlich) unterlässt. Dies sind Fälle von Nachlässigkeit und Fahrlässigkeit. In diesen Fällen ist es schwer, ein geeignetes Ereignis zu finden, das der Unterlassung entsprechen sollte, außer man arbeitet mit der Existenz sogenannter negativer Ereignisse, die jedoch metaphysisch umstritten sind.<sup>451</sup>

b) Gefährdungen: Handlungen können andere gefährden, ohne dass tatsächlich ein Schaden für diese Personen eintritt. Wer meint, dass solche Handlungen gegen normative Forderungen ver-

---

<sup>450</sup> Vgl. auch Cane 2002, ch. 4. Im Zusammenhang mit legaler historischer (retrospektiver) Verantwortung hält er fest: „Causal responsibility is not a precondition of all forms of historic legal responsibility. In the legal context, the main role of causation is to trace a link between conduct and outcomes [...]. Historic legal responsibility may attach to conduct as such regardless of its outcome.“ (Cane 2002, 36)

<sup>451</sup> Vgl. Bach 2010.

stoßen, z.B. unmoralisch sind, ist auch interessiert an der Verantwortung für diese Handlungen. In diesen Fällen sind nicht die tatsächlichen Folgen entscheidend, sondern diejenigen, die hätten eintreten können. Eine Entgegnung wäre, dass hier dennoch eine kausale Verbindung besteht, aber keine realisierte, sondern eine kontrafaktische, welche als Grundlage der Verantwortungszuschreibung fungiert. Die kausale Beziehung zwischen Subjekt, Handlung und Handlungsfolgen wird dann nicht als tatsächliche Verursachungsbeziehung, sondern als mögliche Verursachungsbeziehung gedeutet.

c) Versuche: Versuche sind gescheiterte Handlungen, für die Verantwortung zugesprochen wird, obwohl sie nicht die intendierten Resultate erzielen. Auch hier zählt offenbar nicht das tatsächliche Vorliegen einer Verursachungsbeziehung, sondern ein kontrafaktisches: die Handlung wäre prinzipiell dazu geeignet gewesen, das Resultat zu erzielen. Man könnte meinen, dass in diesen Fällen eben die moralische Norm direkt die Ausführung der Versuchshandlung verbietet statt der Ausführung einer Handlung mit bestimmten Folgen. Diese disjunktive Lösung ist möglich, doch nicht sonderlich attraktiv, da die Begründung für das Verbot von Versuchen damit völlig verschieden wäre von der Begründung für das Verbot der erfolgreichen Handlung. Utilitaristische Theorien haben generell ein Problem mit der Begründung von negativen Sanktionen angesichts von Schäden, die nicht eintreten. Deontologische Theorien sind diesen Fällen leichter zugänglich, wenn man annimmt, dass Rechte und Pflichten immer mit bestimmten Handlungstypen verbunden sind, die nur durch typische Zweck-Mittel-Verhältnisse individuierbar sind. In jedem Fall gilt, dass hier offenbar kein singuläres Kausalurteil über eine faktisch stattgefundene Verursachung als notwendige Bedingung für die Zuschreibung von Verantwortung vorliegt, sondern ein kontrafaktisches Urteil über eine mögliche Kausalbeziehung.

Das Kriterium der Verursachung sollte demnach nicht als tatsächlich vorliegende Verursachungsbeziehung zwischen konkreten Ereignissen gedacht werden, sondern als prinzipielle Möglichkeit der kausalen Verursachung des jeweiligen Handlungsresultates. In Fällen, in denen die Person keinen kausalen Einfluss auf ein Geschehen nehmen kann, unterbleibt normalerweise auch die Zuschreibung von Verantwortung. Ausnahmen von dieser Regel betreffen z.B. Handlungen, die dauerhafte Gefährdungen darstellen, wie z.B. der Betrieb eines Autos. Für diese wird Verantwortung auch ohne den konkreten Nachweis einer kausalen Beziehung zugeschrieben, d.h. auch wenn in konkreten Fällen die Person keine Möglichkeit hatte, den Schaden zu verhindern. Es reicht hier der Nachweis, dass das Resultat ein typisches Ergebnis der betreffenden gefährdenden Handlung ist. Die klassische Lösung, mit diesem Umstand umzugehen, ist die Einführung

entsprechender Versicherungssysteme.<sup>452</sup> Dieser Umstand könnte auch als reine Beweislastumkehr interpretiert werden: derjenige, der eine gefährliche Handlung setzt, z.B. ein Kraftwerk betreibt, der muss im Falle eines Unglückes zeigen, dass das Betreiben der Anlage nicht die Ursache für die aufgetretenen Schäden war.

Die epistemische Situation der Person ist ebenso ein entscheidendes Moment für die Zuschreibung von Verantwortung. Diese betrifft sowohl das tatsächlich vorhandene praktische Wissen der Person als auch dasjenige praktische Wissen, das ihr zugemutet werden kann:

a) Tatsächlich vorliegendes praktisches Wissen einer Person verstehe ich hier in einem sehr weiten Sinn: es umfasst das Wissen um mögliche Handlungstypen und deren kanonische Verfahren, sie zu realisieren, um die jeweilige Handlungssituation, um kausale Folgen von Handlungen unter den gegebenen Umständen, um soziale Konventionen, normative Standards, Institutionen und soziale Rollen und mit ihnen verbundene Erwartungen. Eine Person, die über das relevante praktische Wissen nicht verfügt, wird nicht verantwortlich gemacht für ihre Handlungen und Unterlassungen sowie deren negative Folgen, es sei denn, dass ihr das jeweilige praktische Wissen zugemutet werden kann. Wichtig ist, dass durch diesen Bezug auf praktisches Wissen die Zuschreibung von Verantwortung prinzipiell dynamisch ist und sich ständig verändert. Wenn sich das praktische Wissen einer Person vergrößert, z.B. durch die Erschließung neuer Handlungsmöglichkeiten, so vergrößert sich normalerweise auch die Reichweite der Verantwortungszuschreibungen: die Person wird für Handlungen und Unterlassungen verantwortlich gemacht, für die sie zuvor nicht verantwortlich gewesen wäre.

b) Zumutbares praktisches Wissen: Darunter verstehe ich Wissen um Gegebenheiten, die unter a) aufgeführt sind, nur dass die Person aktuell nicht über dieses Wissen verfügt hat, als sie handelte. Entscheidend ist, dass ihr dieses Wissen jedoch zumutbar war, und die Person daher verantwortlich ist für die entsprechenden Handlungen und Handlungsfolgen. Die genauen Kriterien, wann welches Wissen zumutbar ist, sind ebenfalls dynamisch und kontextuell auf die jeweilige Gesellschaft zugeschnitten. In einer sogenannten Wissensgesellschaft ist der Standard an zumutbarem praktischen Wissen wesentlich höher als in anders strukturierten Gesellschaften.

Was bedeuten nun diese verschiedenen Kriterien angewendet auf Willensschwäche? Ohne die genaue Kenntnis der willensschwachen Handlung, ihrer Umstände, der konkreten Handlungsfolgen und des relevanten Normhintergrundes lassen sich das genaue Maß und die Art von Verantwortung nicht feststellen. Ich kläre hier demnach nur, ob *normalerweise* die Kriterien für

---

<sup>452</sup> Versicherungen sind generell ein gutes Beispiel dafür, dass jemand die Verantwortung für Handlungsfolgen übernehmen kann, ohne kausalen Einfluss auf das Verhalten zu haben: die Versicherung übernimmt (bis zu einem gewissen Grad) die Haftung und ist verantwortlich für die Kompensation des Geschädigten.

eine erfolgreiche Zuschreibung von Verantwortung vorliegen. Willensschwache Handlungen sind intentional. Für die Feststellung der Intentionalität der willensschwachen Handlung reicht es nicht aus, die mentalen Zustände der Person zu kennen. Es muss außerdem anhand des verfügbaren praktischen Wissens innerhalb der jeweiligen Gesellschaft und angesichts der konkreten kausalen Umstände der Handlung und des Handlungsverlaufs entschieden werden, ob hier eine intentionale Handlung vorlag. Die willensschwache Handlung und deren Folgen stehen außerdem im kausalen Einflussbereich der Person. Der genaue Grad der Verantwortung wird grundlegend von dem jeweiligen Handlungstyp beeinflusst, den die willensschwache Person realisiert: diese bestimmt den Normhintergrund als Grundlage für die Verantwortungszuschreibung. Außerdem hängt das konkrete Ausmaß an Verantwortung an den konkreten kausalen Zusammenhängen und tatsächlich eingetretenen Resultaten der Handlung. So ist die willensschwache Person nicht verpflichtet, für Schäden zu haften, die gar nicht eingetreten sind, aber eintreten hätten können. Der erfolgreiche Ausgang der Handlung bestimmt demnach unter anderem das Maß der Verantwortung, die der Person zugeschrieben wird. Die willensschwache Person verfügt im Fall motivationaler Willensschwäche über das notwendige praktische Wissen, vorausgesetzt, ihr besseres Urteil ist wahr oder zumindest gerechtfertigt. Irrt sich die willensschwache Person darüber, was am besten zu tun wäre aus ihrer Sicht, und handelt dann willensschwach gegen ihr eigenes Urteil, kann dies aus Sicht der normativen Bewertung positiv sein: wenn die Handlung, die sie realisiert, also ihre willensschwache Handlung, gut ist. Im Falle exekutiver Willensschwäche muss festgestellt werden, über welches praktische Wissen die Person verfügt hat und ob sich daraus eventuell entschuldigende Umstände ergeben. Für die Frage nach der Verantwortung der willensschwachen Person bleibt noch die Frage zu klären, ob Willensschwäche eine gute Entschuldigung darstellt, weil die Person dabei unter Zwang handle. Ich denke, dass Willensschwäche zu Recht nicht als Entschuldigung akzeptiert wird, aber als mildernder Umstand hinsichtlich von Sanktionen fungiert, die gegenüber der willensschwachen Person verhängt werden. Willensschwäche anzuführen gleicht richtig verstanden daher dem Vorbringen mildernder Umstände und nicht einer Verteidigung auf Unzurechnungsfähigkeit.

### **V.3 Willensschwäche und Zwang**

Gary Watson diskutiert in seinem Artikel *Scepticism about weakness of will* (1977) die Frage, wie Willensschwäche, verstanden als intentionale Handlung wider das eigene, bessere Urteil, möglich sein kann. Dabei interessiert ihn vor allem, wie solche Handlungen zu unterscheiden sind von Handlungen aus Zwang. In einem ersten Schritt kritisiert er die sokratisch-platonische Position, welche motivationale Willensschwäche als Handeln wider das bessere Urteil als leeres Konzept ausweist. Er stimmt der späteren Position von Platon zu, dass die Quellen der Motivation für

Handlungen differenziert sind und nicht alle Strebungen und Wünsche auf das aus der Sicht der Person Gute gerichtet seien.<sup>453</sup>

„Plato’s distinction between the rational and nonrational parts of the soul may be taken as a distinction between sources of motivation. The rational part of the soul is the source of evaluations – judgments as to the value or worth of a particular course of action or state of affairs. The nonrational part of the soul is the source of such desires as arise from appetites and emotions. These desires are blind in the sense that they do not depend fundamentally upon the agent’s view of the good.” (Watson 1977, 320)

Auf diese Weise ist einsichtig, wie eine Person motiviert sein kann, wider ihr besseres Urteil zu handeln. Es bleibt jedoch das Problem, wie diese Handlungen als frei zu verstehen sind, oder genauer, wie sich diese Handlungen von Zwangshandlungen unterscheiden, denn nicht alle Handlungen wider das bessere Urteil sind willensschwach. Watson kritisiert eine übliche Art, zwischen Zwang und Willensschwäche zu unterscheiden.<sup>454</sup> In beiden Fällen sei das, was die Personen tun, nicht die Handlung, die sie evaluativ als beste auszeichnen. Im Unterschied zu willensschwachen Handlungen aber habe die Person bei zwanghaften Handlungen nicht anders handeln können, als sie es getan hat. Dabei wird Willensschwäche zumindest teilweise als Erklärung für das Verhalten der Person gebraucht.

„[...] [W]e are inclined to contrast weakness and compulsion like so: in the case of compulsive acts, it is not so much that the will is too weak as that the contrary motivation is too strong; whereas, in weakness of will properly so-called, it is not that the contrary motivation is too strong, but that the will is too weak.” (Watson 1977, 327)

Das Problem ist nun, wie zwischen den beiden Fällen zu unterscheiden ist, da der Begriff der motivationalen Stärke oder der Stärke eines Wunsches oder einer Strebung obskur sei. Wenn die Stärke eines Wunsches relativ wäre und sich nur über das tatsächliche Handeln einer Person feststellen ließe, dann unterschiede sich Willensschwäche nicht von Zwang, da in beiden Fällen die relevante Strebung bzw. der Wunsch stärker wäre als die alternativen Strebungen bzw. Wünsche. Watson charakterisiert das übliche Konzept der Zuschreibung von Verantwortung für Willensschwäche folgendermaßen: die willensschwache Person verfügte über die Fähigkeit, den auftretenden Strebungen zu widerstehen und ihrem besseren Urteil zu folgen, jedoch hat sie diese Fähigkeit nicht ausgeübt. Das Problem an dieser Theorie ist, dass nicht klar ist, weshalb die Person ihre Fähigkeit nicht eingesetzt hat. Die Antwort könne nach Watson nicht lauten, dass die Strebung in diesem Fall so stark war, da im Falle einer willensstarken Person eine Strebung mit gleicher Stärke auftreten könne, ohne dass die Person ihr nachgeben würde.<sup>455</sup> In jedem Fall bindet diese Theorie die Zuschreibung von Verantwortung an die zum Zeitpunkt der Handlung tatsächlich vorliegende Möglichkeit der Person, anders zu handeln, als sie es getan hat. Dagegen

---

<sup>453</sup> Diese Unterscheidung verwendet Watson auch als Basis seiner Kritik an Davidsons beiden Prinzipien P1 und P2, indem er zwischen zwei Bedeutungen von ‚etwas als besser beurteilen‘ und ‚etwas mehr wollen‘ unterscheidet.

<sup>454</sup> Vgl. Watson 1977, 326ff.

<sup>455</sup> Vgl. Watson 1977, 330.

entwickelt Watson ein Konzept der Zuschreibung von Verantwortung für Willensschwäche, das eine kontrafaktische Bedingung formuliert und sich an der Verantwortungszuschreibung für fahrlässige Handlungen orientiert.<sup>456</sup> Die willensschwache Person ist verantwortlich, da sie mittels eines normalen Maßes an Selbstkontrolle ihren Strebungen für die willensschwache Handlung hätte widerstehen können. In Analogie: eine fahrlässig handelnde Person ist verantwortlich, da sie mit einem normalen Maß an Kontrolle und Aufmerksamkeit den Fehler bzw. die negativen Konsequenzen ihrer Handlungen hätte verhindern können.

„In summary, then, there are capacities and skills of resistance which are generally acquired in the normal course of socialization and practice, and which we hold one another responsible for acquiring and maintaining. Weak agents fall short of standards of “reasonable and normal” self-control (for which we hold them responsible), whereas compulsive agents are motivated by desires which they could not resist even if they met those standards.” (Watson 1977, 331f.)

Zwanghafte Handlungen unterscheiden sich demnach von willensschwachen Handlungen dadurch, dass sie den kontrafaktischen Test nicht erfüllen: selbst unter Anwendung eines normalen Maßes an Selbstkontrolle könnte die Person ihnen nicht widerstehen.<sup>457</sup> Entscheidend ist die Bindung der kontrafaktischen Bedingung an einen kontextuell bestimmbar Standard. Gesellschaften und Gruppen können sich darin unterscheiden, welchen Standard an Selbstkontrolle sie als normal voraussetzen. Was willensschwach in der einen Gruppe ist, könnte als zwanghaft in einer anderen angesehen werden.

„One obvious point is that, on this account, a desire is compulsive (irresistible) or not only relative to certain norms of self-control; only relative to the capacities of the normal person, and this probably means the typical adult in our society.” (Watson 1977, 332)

Die Lösung von Watson halte ich generell für richtig, doch möchte ich folgende Punkte noch nachschärfen:

a) Das Modell legt nahe, dass die willensschwache Person nur indirekt für ihre willensschwache Handlung verantwortlich ist und direkt dafür, dass sie nicht die nötigen Standards an Selbstkontrolle erworben hat. Dagegen meine ich, dass die willensschwache Person direkt für ihre willensschwache Handlung verantwortlich ist, lediglich die Form der Verantwortung ist in Analogie zu derjenigen für fahrlässige Handlungen. Der entscheidende Punkt des kontrafaktischen Konzeptes ist es, nur die Frage zu klären, ob eine Person mit einem normalen Maß an Selbstkontrolle einem Typus an Strebungen hätte widerstehen können oder nicht,

---

<sup>456</sup> „Morally, the proposed account likens weakness to negligence in its emphasis on the point that blameworthiness does not require that one be able to resist or that weak behavior be fully under voluntary control. For negligence is a paradigm case of blameworthy but nonvoluntary behavior.” (Watson 1977, 334)

<sup>457</sup> Vgl. auch Audi 1990 und der Verweis auf ein normales Level an Willenskraft relativ zu einer Situation: „Nonetheless, where an agent is genuinely compelled – which, to be sure, may imply that a person of normal will power relative to the situation would also be unable to forbear – the action is not incontinent.” (Audi 1990, 272)

unabhängig davon, ob die konkrete willensschwache Person diese Fähigkeit zur Selbstbeherrschung erworben hat oder nicht.

b) Der kontrafaktische Test sollte so verstanden werden, dass er über das tatsächliche Vorliegen eines normalen Maßes an Selbstkontrolle der willensschwachen Person nichts impliziert. Dagegen meint Watson, dass nach seinem Modell die willensschwache Person *nicht* über ein normales Maß an Selbstkontrolle verfügt und dennoch verantwortlich ist.<sup>458</sup> Dies ergebe sich daraus, dass im Fall von Willensschwäche keine adäquate Erklärung vorhanden sei, weshalb die willensschwache Person ihre Fähigkeit zur Selbstkontrolle nicht ausgeübt hat. Eine solche Erklärung verweise normalerweise auf a) fehlende Kenntnis der Umstände, b) fehlendes Wollen bzw. fehlende Entscheidung oder c) fehlende Fähigkeit. Da die willensschwache Person gemäß Definition über das relevante praktische Wissen verfügt, kann a) keine Erklärung für ihr Verhalten darstellen.<sup>459</sup> Auch stehe Erklärung c) nicht offen, da es gerade um die Frage geht, ob es einen Sinn macht das Verhalten der Person trotz vorhandener Fähigkeit zu erklären.

Innerhalb von b) verbleiben zwei Möglichkeiten: die Person habe sich erstens nicht für die bessere Handlung entschieden, oder sie habe sich zweitens nicht genügend angestrengt. *Erstens*: eine fehlende Entscheidung könne keine Erklärung für das relevante Handeln sein, da der Begriff der Entscheidung eine Anwendung der relevanten Normen und Werte aus Sicht der handelnden Person in der jeweiligen Situation voraussetze:

„For the capacity of self-control involves the capacity to counteract and resist the strength of desires which are contrary to what one has chosen or judged best to do.” (Watson 1977, 336)

Hier wird ohne weiteres Argument die Situation der willensschwachen Person so charakterisiert, dass sie sich bereits für eine Handlung *entschieden* habe, diese dann aber nicht umsetze, und nicht als eine Situation, in der sie eine Handlung *als beste beurteilt* hat, und dann dennoch eine andere Handlung intendiert hat. Zum einen ist ‚Entscheidung‘ ambivalent zwischen der Festlegung auf ein Handeln in einem Urteil und der Festlegung auf eine Handlung durch eine Intention.<sup>460</sup> Zum anderen widerspricht dies Watsons eigener Differenzierung zwischen *Evaluation von* und *Motivation für* Handlungen. Wenn nämlich zwischen besseren Urteilen und Handlungsintentionen keine notwendige Verknüpfung besteht, dann ist es möglich, die willensschwache Handlung so zu erklären, dass die Person sich nicht für die als besser beurteilte Handlung entschieden hat, und

---

<sup>458</sup> Vgl. Watson 1977, 335ff.

<sup>459</sup> „To return to the case of the weak drinker, the woman judges that she should resist and therefore, by her own lights, has sufficient reason for exercising her alleged capacity. What might explain her not doing so? There seem to me only two possible explanations. (1) She chooses not to. (2) Her effort to resist is culpably insufficient. Both of these explanations will be found inadequate.” (Watson 1977, 336)

<sup>460</sup> „I’ll avoid talk of ‘decisions,’ since ‘to decide’ is ambiguous between the judgment side and the intention or choice side. You can decide that you should  $\varphi$ , a practical judgment, and you can decide to  $\varphi$ , a choice.” (Hinchman 2009, 423, fn. 2)

ebenso, dass sie sich angesichts von vorhandenen Wünschen für die willensschwache Alternative und gegen die Ausübung von Selbstkontrolle entschieden hat. Watson kritisiert, dass diese Lösung die willensschwache Person der rücksichtslosen Person angleiche. Doch dafür besteht kein Grund: beide entscheiden sich für die jeweilige Handlung, aber die willensschwache Person wider ihr besseres Urteil, die rücksichtslose im Einklang mit ihrem besseren Urteil.

*Zweitens:* ungenügende Anstrengung sei in diesem Fall keine mögliche Erklärung, da eine weitere Erklärung dafür nötig sei, weshalb die willensschwache Person sich nicht mehr anstrengt, eine solche aber nicht gegeben werden könne. Erklärungen für ungenügende Anstrengung verwiesen nämlich entweder auf die fehlerhafte Einschätzung des nötigen Aufwandes für die Handlung oder auf ein Urteils, dass sich ein gewisser Aufwand nicht lohne. Beide Möglichkeiten seien aber ausgeschlossen bei Willensschwäche: eine Fehleinschätzung wäre keine Willensschwäche, sondern ein anderer kognitiver Fehler,<sup>461</sup> und die willensschwache Person verfügt über ein bestehendes besseres Urteil, das den erforderlichen Aufwand für die Anstrengung rechtfertigt. Ich stimme Watson zu, dass die Erklärung aus fehlender Anstrengung im Grund nur dann funktioniert, wenn man bereits die Erklärung aus fehlendem Entschluss akzeptiert. Sie kann jedoch nützlich sein um zwischen Fällen zu differenzieren, in denen jemand Maßnahmen zur Selbstkontrolle gesetzt hat, und dennoch nicht widerstanden hat, und Fällen, in denen jemand überhaupt nicht versucht hat, sich selbst zu kontrollieren. Generell ist es problematisch, der willensschwachen Person aufgrund ihrer willensschwachen Handlung die Fähigkeit zur Selbstkontrolle abzusprechen: eine Person könnte in ihrem Leben eine einzige willensschwache Handlung ausüben und ansonsten ständig über ein normales Maß an Selbstkontrolle verfügen.<sup>462</sup> Der Test, ob jemand über ein normales Maß an Selbstkontrolle verfügt, kann nicht daran liegen, ob jemand ständig selbstkontrolliert handelt. In Analogie zu Fällen von fahrlässigem Handeln: die einmalig unaufmerksame Person verfügt in den meisten Fällen über das nötige Maß an Aufmerksamkeit. Für die Frage nach der Verantwortung ist es daher nebensächlich, ob sie zum Zeitpunkt der Handlung über die Fähigkeit zur Selbstkontrolle verfügt hat oder nicht. Entscheidend ist nur, was eine normale Person mit einem normalen Maß an Selbstkontrolle hätte tun können. Es ist also nicht wichtig, ob die Person die Fähigkeit zur Selbstkontrolle nicht hatte, sie aber hätte entwickeln sollen, oder ob sie die Fähigkeit hatte, sie aber nicht eingesetzt bzw. ausgeübt hat. Beides ist möglich, Watson geht jedoch davon aus, dass stets die erste Option zutrifft, die Person demnach nicht über das nötige Maß an Selbstkontrolle verfügt hat. Dies ist identisch zur Zuschreibung von Verantwortung für fahrlässige Handlungen: die Frage nach der Verantwortung für fahrlässige Handlungen ist unabhängig davon, ob die Person hätte anders

---

<sup>461</sup> Vgl. die Diskontierungstheorien von Willensschwäche Kap. III.2.

<sup>462</sup> Vgl. zur Fähigkeit zur Selbstkontrolle auch Kennett & Smith 1994.

handeln können, als sie es getan hat. Entscheidend ist, wie eine normale Person hätte handeln können. Ein Beispiel: Ein Autofahrer vergisst die Handbremse seines Wagens in abfallendem Gelände anzuziehen, woraufhin sich der Wagen in Bewegung versetzt und einen Unfall verursacht. Es ist nicht klar, ob *diese* Person in *dieser* Situation tatsächlich anders hätte handeln können, als sie es getan hat, d.h. sich an das Anziehen der Handbremse zu erinnern und dementsprechend zu handeln. Dieser Umstand ist jedoch irrelevant für die Zuschreibung der Verantwortung für die Verursachung des Unfalles an den Autofahrer. Relevant ist, ob eine dem Autofahrer vergleichbare Person mit einem für diese Gesellschaft typischen Maß an Aufmerksamkeit in solch einer Situation anders hätte handeln können, d.h. sich an das Anziehen der Handbremse zu erinnern. Es ist demnach möglich, dass eine Person über das normale Maß an Selbstkontrolle verfügt und dennoch willensschwach handelt. Die Erklärung für dieses Verhalten wird letztlich ungenügend bleiben, da sie auf den Begriff des Entschlusses einer Person rekurriert, der nicht weiter durch andere Faktoren erklärt wird. Die Entscheidung der willensschwachen Person wird aber verständlich, wenn ihre Situation und ihre in der Situation auftretenden Wünsche berücksichtigt werden, die fähig sind, direkt Strebungen für Handlungen auszulösen, ohne von der Person als gut beurteilt zu werden.

c) Die Frage, ob eine Person über ein gewisses Maß an Selbstkontrolle verfügt und dieses nicht angewendet hat, oder ob sie dieses erst gar nicht erlernt hat, ist nicht für die Zuschreibung von Verantwortung entscheidend, sondern beeinflusst das Ausmaß an Sanktionen, den moralischen Reaktionen (*reactive attitudes*), die wir der Person aussetzen.<sup>463</sup> Dass eine Person ein normales Maß an Selbstkontrolle gar nicht entwickelt hat, und dies vielleicht aus verständlichen Gründen, untergräbt nicht die Zuschreibung von Verantwortung dieser Person für ihre willensschwachen Handlungen, sondern setzt die Schärfe an Sanktionen herab, die wir gegenüber der Person setzen.

d) Die Zuschreibung der Verantwortung der willensschwachen Person für ihre willensschwache Handlung wird nach diesem Modell gerechtfertigt dadurch, dass eine Person mit einem für diese Gesellschaft normalen Maß an Selbstkontrolle den willensschwachen Strebungen hätte widerstehen können. Es bleibt die Frage, ob Personen mit einem normalen Maß an Selbstkontrolle bestimmten Strebungen widerstehen können, und ob diese Tatsache eine Bedingung ist für die Zuschreibung von Verantwortung. Diese Frage ist sicherlich wichtig, doch würde sie in diesem Kontext zu weit führen, da dies metaphysische Fragen der Verantwortung betrifft. Nach einem alltäglichen Verständnis ist es jedoch völlig unproblematisch davon auszugehen, dass wir

---

<sup>463</sup> Zu *reactive attitudes* vgl. den mittlerweile klassischen Beitrag von Peter Strawson 1962.

unsere Strebungen, Emotionen und Wünsche (teilweise) kontrollieren können. Ein großer Teil unserer Erziehungspraxis wäre ansonsten wohl völlig unverständlich.

## **Zusammenfassung**

Der Begriff der Verantwortung ist komplex und besitzt verschiedene Funktionen. Normalerweise erfüllen willensschwache Personen und ihre willensschwachen Handlungen die Kriterien für eine erfolgreiche Zuschreibung von Verantwortung. Philosophisch schwierig ist hier vor allem die Unterscheidung von zwanghaften und willensschwachen Handlungen. Dafür greife ich auf das Konzept von Garry Watson und seiner kontrafaktischen Bedingung der Zuschreibung von Verantwortung zurück: Hätte die willensschwache Person das innerhalb ihrer Gesellschaft erwartete Maß an Selbstkontrolle ausgeübt, hätte sie die willensschwache Handlung vermeiden können. Dies unterscheidet sie von der zwanghaft handelnden Person, die selbst unter Aufbietung dieses Maßes an Selbstkontrolle nicht anders hätte handeln können, als sie es getan hat. Ob die willensschwache Person zum Zeitpunkt des Handelns über die Fähigkeit zur Selbstkontrolle tatsächlich verfügt hat, ist für die Zuschreibung von Verantwortung nicht relevant. Gegen Watson habe ich jedoch argumentiert, dass die Person tatsächlich über die Fähigkeit zur Selbstkontrolle verfügen könnte, selbst wenn sie willensschwach gehandelt hat. Fähigkeiten müssen nicht ständig erfolgreich aktualisiert sein, damit sie zugeschrieben werden können. Außerdem ist es nach einer Trennung der Begriffe von besserem Urteil und Entscheidung, etwas zu tun, möglich, dass eine Person sich für eine Handlung entscheiden kann, die nicht ihrem besseren Urteil entspricht.

## **VI Problemfeld Normativität**

Die Zuschreibung von Verantwortung für Willensschwäche und willensschwache Handlungen macht nur Sinn vor einem normativen Hintergrund, anhand dessen diese Zuschreibung erfolgt. Diese umfassen normative und evaluative Standards, deren Erfüllung, Nichterfüllung und Übererfüllung (Supererogation) als Basis für die Verantwortungszuschreibung fungieren können. Bei willensschwachen Handlungen hängen diese Standards primär davon ab, welcher Handlungstyp ausgeführt wurde, welche kausalen Konsequenzen die Handlung hervorgerufen hat oder hervorrufen hätte können, und welche sozialen oder institutionellen Rollen die Person innehatte. Weiters können willensschwache Handlungen auch spezielle, durch Vereinbarung getroffene Pflichten berühren, welche die Person übernommen hat. In erster Linie ist demnach die Frage nach der evaluativen und normativen Bewertung von willensschwachen Handlungen parasitär gegenüber der normativen und evaluativen Bewertung derselben Handlung abgesehen von ihrem willensschwachen Aspekt: Im Vordergrund steht die Bewertung des tatsächlichen Handelns, die Bewertung der Willensschwäche ist dagegen sekundär. Die Aufgabe hier ist demnach, diesen relativen und indirekten normativen Charakter von Willensschwäche zu analysieren und die verschiedenen normativen Gesichtspunkte willensschwacher Handlungen offenzulegen. Ich beginne mit der Frage, ob und in welchem Sinn willensschwache Handlungen irrational sind (1). Es folgt eine tugendethische Betrachtung von Willensschwäche als Charakterzug und relatives Laster (2). Die Analyse von Willensschwäche als einem Handeln aus mangelndem Selbstvertrauen (3) führt schließlich zur Analyse von Willensschwäche als Untergrabung des Selbstrespekts einer Person (4), einem zentralen Gut im Leben eines jeden handelnden Akteurs. Ich schließe mit einer Diskussion über die Möglichkeit globaler Willensschwäche.

### **VI.1 Willensschwäche als Form praktischer Irrationalität**

Ebenso wie theoretische Rationalität kann praktische Rationalität in einem internen, engen (oder subjektiven) und in einem externen, weiten (oder objektiven) Sinn verstanden werden. Für ein internes Verständnis von Rationalität verlangt diese bloß interne Kohärenz und Konsistenz, die Abwesenheit von Widersprüchen und weiterer Bedingungen für das Bestehen von Kohärenz, während ein externes Konzept Rationalität als adäquate Reaktion auf gegebene Gründe auslegt, wobei diese Gründe als Fakten gedeutet werden, die bestehen oder nicht bestehen.<sup>464</sup> Innerhalb der Diskussion um praktische Rationalität deckt sich dies oft mit der Unterscheidung zwischen einem rein instrumentellen und einem substanziellen Verständnis praktischer Vernunft. In einem

---

<sup>464</sup> Vgl. die Unterscheidung bei Hooker & Streumer 2004 zwischen prozeduraler (Instrumentalismus) und substantieller praktischer Rationalität. Vgl. auch Radoilska 2012

instrumentellen Verständnis ist Zweck-Mittel-Rationalität, die Wahl der adäquaten Mittel für einen bestehenden Zweck, die einzig legitime Form praktischer Rationalität. In einem substanziellen Verständnis sind auch die Zwecke einer Person legitime Gegenstände einer rationalen Kritik.

Zusätzlich zu dieser Unterscheidung kann Rationalität normativ oder rein deskriptiv verstanden werden. Für ein normatives Verständnis von Rationalität sind rationale Standards oder Normen mehr als bloße Beschreibungen dessen, was rationale Akteure glauben oder tun würden, sondern Anforderungen an rationale Akteure, diesen Standards zu entsprechen. Nach einem deskriptiven Verständnis sind Anforderungen der Rationalität selbst nicht normativ, sondern lediglich Beschreibungen dessen, was es heißt, rational zu denken und zu handeln. Für Nomy Arpaly sind dies zwei unterschiedliche Weisen, über Rationalität nachzudenken, bzw. zwei unterschiedliche Ansprüche an eine Theorie der Rationalität. Eine entwickelt ein Manual für Personen, das angibt, was diese tun sollen, um rational zu handeln, die andere entwickelt lediglich eine Theorie, die rationales von irrationalem Verhalten unterscheidet.<sup>465</sup>

„There are at least two ways to think of theorizing about rationality. One way is to see the ideal theory of rationality as providing us with a manual of sorts: follow these instructions, and you will always make a rational decision, or you will at least know whether the course of action that you are taking is rational or not. Another way is to see theorizing about rationality as aiming not at providing us with a manual but at providing us simply with a theory – a theory that tells us when people act rationally and when they do not, so that given a God's-eye view of a person's circumstances, beliefs, and motives, one would be able to tell how rational or irrational said person would be in performing a certain action.“ (Arpaly 2000, 488)

Durch Kombination dieser Unterscheidungen ergeben sich zumindest vier verschiedene Positionen:

	Intern	Extern
Deskriptiv	Rationale Anforderungen sind intern und deskriptiv	Rationale Anforderungen sind (auch) extern und deskriptiv
Normativ	Rationale Anforderungen sind intern und normativ	Rationale Anforderungen sind (auch) extern und normativ

Für Positionen, die Rationalität als deskriptiv ansehen, reicht es für eine normative Bewertung von willensschwachen Handlungen nicht aus zu zeigen, dass diese irrational sind. Dies kann in

<sup>465</sup> Als Analogie dazu aus dem Bereich der theoretischen Rationalität dienen Diskussionen über den Status logischer Gesetze oder Axiome. Nach einer Ansicht untersucht die Logik lediglich deskriptiv den Begriff der Folgerichtigkeit in all seinen Facetten, während nach einer anderen die Logik die Kunst des korrekten Schließens lehrt. Ich denke, dass beide Ansichten richtig sind und einander nicht ausschließen, auch wenn der derzeit übliche logische Formalismus nur ungenügend ist, um zugleich Akte des Schließens als auch Schlüsse (als deren gegenständlicher Seite) wiederzugeben. Freges Behauptungsstrich war ein erster Versuch, zumindest ein formales Zeichen für den Akt des Behauptens einzuführen, der aber nicht integriert ist in den sonstigen logischen Formalismus. Die konstruktive Typentheorie von Per Martin-Löf ist eine moderne Möglichkeit, diesen Mangel zu beheben. Vgl. Martin-Löf 1984; Frege 1993.

einer Gründe-Terminologie so ausgedrückt werden, dass die Irrationalität von willensschwachen Handlungen noch keinen guten (hinreichenden) Grund dafür liefert, diese zu unterlassen.<sup>466</sup> Für normative Auffassungen von Rationalität ist die Kennzeichnung von willensschwachen Handlungen als irrational selbst bereits ein (prima facie) Grund, sie nicht zu tun. Ich werde mich vorwiegend mit Auffassungen beschäftigen, nach denen Standards praktischer Rationalität zumindest als schwach normativ verstanden werden, also prima facie Gründe liefern, den Standards zu entsprechen. Für Positionen, die praktische Rationalität rein deskriptiv verstehen, bieten die weiteren Kapitel evaluative und normative Analysen, weshalb Willensschwäche generell vermieden werden sollte. Zwei Fragen stellen sich auf Basis der vorigen Überlegungen: 1) Ist willensschwaches Handeln notwendigerweise *intern* irrational? 2) Ist willensschwaches Handeln notwendigerweise *extern* irrational?

### **VI.1.1 Interne Irrationalität von Willensschwäche**

Die allermeisten Theorien zu Willensschwäche gehen davon aus, dass willensschwache Handlungen zumindest intern irrational sind. Entscheidend ist im Vorfeld dieser Frage jedoch, welche Art von Willensschwäche hier gemeint ist. Bei synchroner motivationaler Willensschwäche intendiert eine Person eine Handlung, obwohl sie eine ihr mögliche alternative Handlung als besser einschätzt. Die interne Inkonsistenz liege demnach darin, zwei mentale Zustände zu haben, die einander widersprechen. Sie liegt nicht in der klassischen Inkonsistenz, gleichzeitig etwas und dessen Gegenteil zu glauben, da bei motivationaler Willensschwäche ein Glaubenszustand und eine Intention einander widersprechen. Doch wie kann eine Intention einem Glaubenszustand widersprechen? Dies ist dann möglich, wenn das Halten beider mentalen Zustände einer rationalen Anforderung widerspricht. Eine andere Möglichkeit ist, die Irrationalität synchroner motivationaler Willensschwäche nicht als Widerspruch zweier mentaler Zustände zu deuten, sondern als fehlerhaften praktischen Überlegungsprozess: Der Überlegungsprozess stützt nicht die Intention, zu der die Person gelangt ist, in Analogie zu einem theoretischen Fehlschluss, in der die Person zu einer Konklusion gelangt, die nicht aus den Prämissen folgt.

Es lassen sich zumindest drei Arten von praktischer Überlegung unterscheiden: a) Überlegungen darüber, was am besten zu tun ist; b) Überlegungen, die das, was man glaubt tun zu sollen, mit Intentionen verknüpft, es auch zu tun; c) Überlegungen, wie man etwas, was man zu tun

---

<sup>466</sup> Mit Gründe-Terminologie meine ich hier die reduktive Analyse von Normativität in Form von Gründen: Gründe sind nach dieser Position das normative Grundphänomen, auf dem sich die Normativität aller anderen Phänomene, also z.B. von Normen, herleitet. Für eine ältere Auffassung sind Gründe nur deswegen normativ, da sie in Relation stehen zu Normen. Ich denke, für die normative Analyse von Willensschwäche ist dieser Unterschied nicht entscheidend.

intendiert, auch umsetzen kann. Die Unterscheidung zwischen theoretischen und praktischen Überlegungen kann sich entweder am *Inhalt* solcher Überlegungen oder an der *Art* der beteiligten mentalen Zustände bzw. propositionalen Einstellungen (den mentalen Akten) gegenüber den Inhalten orientieren. So sind für einige Positionen vorgeblich praktische Überlegungen der Art a) rein theoretische Überlegungen über etwas Praktisches. Für andere sind praktische Überlegungen im Sinne von a) keine rein theoretischen Überlegungen, sondern bereits Überlegungen der Form b) oder c). Dies sind Positionen, die Urteile darüber, was am besten zu tun ist, als Intentionen deuten, es auch zu tun. So ist der Präskriptivismus eine semantische Theorie moralischer Prädikate, welche rein theoretische Überlegungen über Praktisches aus semantischen Gründen ausschließt.<sup>467</sup> Ich denke, dass Überlegungen, die nur von Glaubenszuständen zu anderen Glaubenszuständen fortschreiten, keine praktischen Überlegungen sind, sondern Überlegungen *über* etwas Praktisches. Ebenso halte ich Überlegungen der Art b) für keine Beispiele gelungener praktischer Überlegungen, da sie nicht folgerichtig sind.

Im Folgenden gebe ich die Form der jeweiligen Überlegung und ein konkretes Beispiel einer solchen Überlegung an, die im Alltag oftmals nur als Enthymem vorliegt, d.h. dass der Person einige der Prämissen nicht explizit gegeben sind. In der Struktur gebe ich explizit die Art der propositionalen Einstellung (Akt) an, gefolgt von der Proposition, auf die sich die jeweilige Einstellung richtet (Akt: Proposition). Der Akt bzw. die Art der Einstellung ist nicht Teil des Inhalts der Proposition. Dies ist analog zur Analyse der Behauptung (Assertion) einer Proposition durch Frege und der Einführung eines Behauptungsstriches zur Kennzeichnung des Sprechaktes bzw. des Modus, in dem ein beurteilbarer Inhalt (der Gedanke bzw. die Proposition) gegeben ist.<sup>468</sup> Diese Schemata an Überlegungen sind hier dafür gedacht, intuitiv verschiedene Möglichkeiten praktischer Überlegungen zu illustrieren, und sind als solche technisch nicht ausgereift oder vollständig.

Überlegungen der Art a) verknüpfen Glaubenszustände mit anderen Glaubenszuständen:

<b>Struktur</b>	<b>Beispiel</b>
Glaube: Jeder in Situation X soll F-en	Jeder soll seinen Freunden helfen
Glaube: Ich befinde mich in Situation X	Mein Freund benötigt Hilfe
Also Glaube: Ich soll F-en	Also soll ich ihm helfen

<sup>467</sup> Vgl. Kap. III.3.

<sup>468</sup> Vgl. in diesem Zusammenhang auch die Anwendung der konstruktiven Typentheorie von Per Martin-Löf auf die Linguistik durch Aarne Ranta in Ranta 1994.

Überlegungen der Form b) verknüpfen Glaubenszustände mit Intentionen, diese zu tun.

Struktur	Beispiel
Glaube: Ich soll F-en	Ich sollte meinem Freund helfen
Glaube: Ich kann F-en	Ich kann meinem Freund helfen
Also Intention: Ich will F-en	Also will ich ihm helfen

Überlegungen der Form c) verknüpfen konative Zustände wie Intentionen mit kognitiven Zuständen wie Glaubenszuständen. Solche praktische Überlegungen sind z.B. Zweck-Mittel-Überlegungen.

Struktur	Beispiel
Intention: Ich will F-en	Ich will ins Kino gehen
Glaube: Wenn ich F-en will, muss ich zuerst G-en	Um ins Kino zu gehen muss ich zuerst eine Karte kaufen
Also Intention: Ich will zuerst G-en	Also will ich zuerst eine Karte kaufen

Die Irrationalität von Willensschwäche könnte darin liegen, nicht dem Schema b) korrekter praktischer Überlegung zu entsprechen. Die Person verfügt über ein Urteil, das den Glaubenszustand ausdrückt, dass sie eine Handlung F tun sollte, und über den Glauben, dass ihr diese Handlung offensteht, dennoch intendiert sie eine andere, bereits in ihrer praktischen Überlegung berücksichtigte Handlung G. Der Gang der Überlegung könnte so ausgedrückt werden:

Struktur	Beispiel
Glaube: Alles in allem ist F-en besser als G-en	Alles in allem ist es besser heute zu arbeiten statt an den Strand zu gehen
Glaube: Also soll ich F-en	Ich sollte heute arbeiten gehen
Glaube: Ich kann F-en	Ich kann heute arbeiten gehen
Also Intention: Ich will aber G-en	Ich will aber an den Strand gehen

In diesem Schema wird jedoch der für Willensschwäche typische Konflikt zwischen zwei einander ausschließenden Handlungen nicht repräsentiert. Grundlegender ist noch das Problem an dieser Art, willensschwache und überhaupt praktische Überlegungen zu verstehen, dass die dabei involvierten Inferenzen nicht wahrheitsbewahrend fungieren.<sup>469</sup> Der praktische ‚Schluss‘

<sup>469</sup> Im Folgenden stütze ich mich wesentlich auf Gedanken von Olav Gjelsvik, die er auf dem Workshop *The nature of the enkratic requirement of rationality*, Wien 26.–27.– April 2013, vorgetragen hat.

nach dem Schema b) ist nicht folgerichtig, da die Konklusion einen Term enthält („wollen“), der in den Prämissen überhaupt nicht vorkommt. Natürlich könnte man für diese Art von Schluss eine eigene Inferenzregel einführen, welche diese rechtfertigt. Ein konservativerer Ansatz, der mit den Schlussregeln auskommt, die bereits für theoretische Überlegungen akzeptiert sind, ist jedoch vorzuziehen.

Dagegen ist für John Broome eine Überlegung praktisch, wenn sie zu einer Intention als Konklusion führt.<sup>470</sup> Sowohl Überlegungen der Art c) als auch solche der Art b) sind damit korrekt. Dafür formuliert er eine normative Anforderung (rational requirement), ein *requirement of enkrasia*, welches einen Glauben, was man tun soll, mit Intentionen, dieses zu tun, verbindet. Wichtig ist, dass diese Anforderung hypothetisch ist, d.h. sich über das gesamte Konditional erstreckt (wide scope interpretation).<sup>471</sup>

Die Rationalität verlangt, dass (wenn X glaubt, dass X F-en soll, dann intendiert X zu F-en)

Das enkratische Prinzip in seiner weiten Form verlangt demnach nur die Kombination gewisser mentaler Zustände, und nicht die Bildung einer Intention oder die Aufgabe eines bestimmten Glaubens. So stellt Broome sicher, dass Rationalität nicht direkt verlangt, eine Handlung zu intendieren, die auf einem falschen normativen Glauben, diese Handlung tun zu sollen, basiert. Trotz dieses umfassenden Konzeptes über die Natur rationaler Anforderungen bleibt die Frage, weshalb diese Art der praktischen Inferenz ein gültiger Schluss sein sollte. Um diesen Schluss folgerichtig zu machen, bedarf es meiner Ansicht nach einer praktischen Prämisse, also einer Prämisse, deren Proposition in einer praktischen Auffassung (in einem praktischen Akt) durch die Person gegeben ist, also z.B. in einer Intention. Ansonsten handelt der Schluss lediglich von etwas Praktischem, anstatt selbst praktisch zu sein.

Struktur	Beispiel
Glaube: Ich soll F-en	Ich soll heute arbeiten
Glaube: Ich kann F-en	Ich kann heute arbeiten
Intention: Wenn ich F-en soll, dann will ich F-en	Wenn ich heute arbeiten soll, dann will ich auch heute arbeiten
Also Intention: Ich will F-en	Also will ich heute arbeiten

<sup>470</sup> „I therefore identify practical reasoning as reasoning that concludes in an intention, which is a mental state.“ (Broome 2001, 175)

<sup>471</sup> Zum Vergleich dazu die Formulierung mit engem Skopus (narrow scope): Wenn X glaubt, dass X F-en soll, dann verlangt die Rationalität, dass (X zu F-en intendiert). Wichtigster Einwand gegen die Interpretation rationaler Prinzipien im Sinne des engen Skopus ist der sogenannte Bootstrapping-Einwand (bootstrapping objection): Eine Person erlangt bloß dadurch, dass sie glaubt, etwas tun zu sollen, bereits einen echten Grund, dies auch zu tun. Dagegen kann die Person nach der weiten Interpretation keinen solchen Grund durch bloßen Glauben herbeiführen.

Das Problem der willensschwachen Person ist nach diesem Modell, dass ihr die entsprechende praktische Prämisse fehlt. Sie gelangt zu einem besseren Urteil über ihre Handlungsoptionen, doch fehlt ihr die entsprechende Intention, auch so zu handeln, wie sie glaubt, dass sie handeln sollte. Stattdessen bildet sie eine Intention im Einklang mit ihren Streben, die sie für die willensschwache Handlung besitzt. Positionen, welche synchrone motivationale Willensschwäche ablehnen, können die Irrationalität von willensschwachen Handlungen nicht auf diese Weise erläutern, da nach diesen die entsprechenden praktischen Überlegungsprozesse anders strukturiert sind. Es gibt innerhalb dieser Theorien keinen Bedarf an praktischen Überlegungen, welche Glaubenszustände darüber, was jemand tun sollte, mit Intentionen verbinden, da bestimmte evaluative Urteile in Wahrheit schon Intentionen sind. Worin liegt in diesen Modellen die Irrationalität verstanden als interne Inkohärenz mentaler Zustände der Person? Sie kann diachron gedeutet werden als unbegründeter Wechsel von Intentionen bzw. als unbegründeter Wechsel von Urteilen darüber, was man am besten tun sollte. Sie könnte auch im Sinne einer Selbsttäuschungstheorie gedeutet werden als das gleichzeitige Glauben einander widersprechender Propositionen über die eigenen Wünsche und Intentionen. Im ersten Fall ist es schwierig, diese Standards tatsächlich als Standards von Rationalität in einem internen Sinn zu deuten und nicht als Standards klugen oder moralisch guten Handelns. Im zweiten Fall hat die Person keinen epistemischen Zugang, ihre eigene Irrationalität zu bemerken, da die Selbsttäuschung dies unterdrückt. Für solche Positionen ist es demnach ungleich schwieriger zu zeigen, dass willensschwaches Handeln tatsächlich irrational ist, da sie davon ausgehen, dass Intention und besseres Urteil miteinander identisch sind und daher die Person immer das tut, was sie selbst als beste Handlung ansieht, wenn sie dies auch tun kann.

Folgend fasse ich die drei Möglichkeiten, die praktische Überlegung des Willensschwachen zu konzeptualisieren, zusammen.

**Skepsis gegenüber synchroner motivationaler Willensschwäche:**

Annahme: Normative Urteile sind motivational, drücken keine reinen Glaubenszustände aus – das bessere Urteil ist zugleich eine Intention, entsprechend zu handeln.

<b>Willensschwäche</b>	<b>Willensstärke</b>
Intention: Alles in allem soll ich F-en statt G-en	Intention: Alles in allem soll ich F-en statt G-en
Intention: Alles in allem will ich aber G-en	
Also Intention: Ich will G-en	Also Intention: Ich will F-en

### Willensschwäche als falscher praktischer Schluss:

Annahme: Es gibt korrekte praktische Schlüsse von reinen Glaubenszuständen zu Intentionen

Willensschwäche	Willensstärke
Glaube: Ich soll alles in allem F-en	Glaube: Ich soll alles in allem F-en
Glaube: Ich kann F-en	Glaube: Ich kann F-en
Also Intention: Ich will aber G-en	Also Intention: Ich will F-en

### Willensschwäche als fehlende praktische Prämisse:

Der willensschwachen Person mangelt es an der korrekten Einstellung gegenüber ihren Glaubenszuständen: sie will nicht tun, was sie selbst glaubt tun zu sollen.

Willensschwäche	Willensstärke
	Intention: Ich will tun, was ich tun soll
Glaube: Ich soll alles in allem F-en statt G-en	Glaube: Ich soll alles in allem F-en statt G-en
Glaube: Ich kann F-en	Glaube: Ich kann F-en
Intention: Ich will aber G-en	Also Intention: Ich will F-en

Es bleibt noch das Problem, dass in diesem Modell der praktischen Überlegung der willensschwachen Person der typische Konflikt, in dem sich die willensschwache Person befindet, nicht abgebildet ist.<sup>472</sup> Weshalb formt die willensschwache Person gerade eine von ihrem präferierten Urteil abweichende Intention? Es fehlen also in diesem Modell die Faktoren, welche aus Sicht der Person für die Ausführung der willensschwachen Handlung sprechen. Die derart ergänzte Struktur der praktischen Überlegung kann folgendermaßen aussehen:

Willensschwäche	Willensstärke
Wunsch: G-en	Wunsch: G-en
Glaube: Ich kann entweder F-en oder G-en	Glaube: Ich kann entweder F-en oder G-en
Glaube: Ich soll alles in allem F-en statt G-en	Glaube: Ich soll alles in allem F-en statt G-en
	Intention: Ich will tun, was ich tun soll
Intention: Ich will aber G-en	Also Intention: Ich will F-en

Die Debatte um das rechte Verständnis praktischer Überlegung ist selbstverständlich umfangreich und wird auf verschiedenen Ebenen geführt. Das hier favorisierte Modell kann mit einiger

<sup>472</sup> Vgl. zu diesem Problem Davidson 1969a, 33ff.

Plausibilität an Vorstellungen von Aristoteles und, vielleicht etwas überraschend, auch Immanuel Kant anschließen. *Erstens* ist das Modell äußerst konservativ, was die Akzeptanz von folgerichtigen Inferenzen betrifft. Nur solche Inferenzen werden als folgerichtig akzeptiert, welche auch ihren Widerpart in rein theoretischen Überlegungen finden. *Zweitens* interpretiert es genuin praktische Überlegungen als Überlegungen, die von zumindest einer in einer praktischen Einstellung, also in einem konativen Akt gehaltenen Proposition, ihren Ausgangspunkt haben, und in einer anderen konativen Einstellung münden. Es braucht eine *praktische* Prämisse, um zu einem *praktischen* Schluss zu gelangen. Dies ist in einer Linie mit den spärlichen Beschreibungen praktischer Überlegungen bei Aristoteles.<sup>473</sup> *Drittens* bleibt es einem Modell von Handlung treu, welches Handlungen als aus zumindest zwei relativ voneinander unabhängigen Elementen, einem konativen und einem kognitiven Element, erklärt. Weder reine kognitive Zustände wie Glaubensakte noch rein konative Zustände wie Wünsche oder Strebungen sind hinreichend für Handlungen, erst in ihrem Zusammenspiel entfalten sie ihre Wirkung. Dies ist auch die Sicht moderner Entscheidungstheorie.<sup>474</sup> *Viertens* braucht dieses Modell praktischer Überlegung nicht humanisch interpretiert zu werden: nicht alle praktische Prämissen müssen auf kontingent auftretenden Wünschen oder Präferenzen basieren. Auch eine Interpretation nach Kant ist hier möglich: Vernunft ist praktisch, indem sie sich selbst ein praktisches Gesetz gibt, das Moralgesetz. Die praktische Prämisse einer praktischen Überlegung kann so als Maxime gedeutet werden, als Willensbestimmung. Die Pointe von Kants praktischer Philosophie ist, dass die Vernunft praktisch wirksam sein kann in Form eines Willens, und nicht, dass rein theoretische Überlegungen durch Vernunftschlüsse in sogenannte ‚praktische‘ Urteile münden, welche dann selbst Intentionen sind oder solche bewirken. In diesem Sinne deute ich Kant als motivationalen Externalisten.<sup>475</sup> *Fünftens* bewahrt dieses Modell eine echte Asymmetrie zwischen theoretischen und praktischen Überlegungen: praktische Überlegungen sind praktisch, da in ihnen vernünftig von bestimmten praktischen Einstellungen auf andere geschlossen wird, und nicht, weil sie von Prämissen mit praktischen Inhalten auf andere Prämissen mit praktischen Inhalten schließen.

### **VI.1.2 Externe Irrationalität von Willensschwäche**

Es ist umstritten, ob Willensschwäche auch in einem externen Sinn notwendigerweise irrational ist. Einige haben Argumente vorgelegt, die nahelegen, dass willensschwaches Handeln rational sein kann im Sinne eines weiten oder externen Verständnisses von Rationalität als adäquate Reaktion auf tatsächlich bestehende Gründe eines Handelnden.<sup>476</sup> Bei Robert Audi ist eine

---

<sup>473</sup> Vgl. Kap. II.1.

<sup>474</sup> Vgl. Heath 2003.

<sup>475</sup> Vgl. Ameriks 2006.

<sup>476</sup> Vgl. Audi 1990; Jones 2003; Arpaly 2000; McIntyre 1990; Döring 2009.

Handlung dann rational, wenn sie mit der Totalität an Gründen übereinstimmt.<sup>477</sup> Da es sein kann, dass diese Gründe der handelnden Person nur partiell gegeben sind, und ein besseres Urteil der Person daher falsch oder irrational sein könnte, könnten Handlungen wider das eigene bessere Urteil in manchen Fällen rational sein.<sup>478</sup>

„Indeed, a practical judgment may itself be passionate, foolish, or even incontinent, in the sense of going against one’s assessment of what the total evidence bearing on one’s options indicates one should do. However, even where the judgment is warranted by the premises on which it based, they themselves may be defective from the overall point of view central to the agent’s rationality, and in that case an incontinent action may turn out to be rational.” (Audi 1990, 280)

Als Beispiel beschreibt Robert Audi den Fall eines Vaters, der überzeugt ist von der Wichtigkeit von Strafe als Vergeltung von Schuld (Vergeltungstheorie), der aber entgegen seinem besseren Urteil davon absieht, seine Tochter für zu langes Telefonieren zu bestrafen und sich daraufhin selbst wegen seiner Willensschwäche kritisiert. Zu einem Fall von inverser Akrasia bzw. Willensschwäche wird dies, indem Audi zudem annimmt, dass der Vater nach längerer Überlegung zum gegenteiligen Urteil gekommen wäre, also seine Tochter nicht zu bestrafen, und dieses Urteil besser sei, da es mehr mit den sonstigen Charaktereigenschaften des Vaters und der Gesamtheit seiner Handlungsgründe harmonieren würde. Deswegen handle der Vater in diesem Fall zwar willensschwach, doch nicht irrational.

„But even though his incontinent action accords with a civilised and generally admirable compassionate desire, it *does* go against his standing better judgment and its underlying retributive desires. Still, the overall rational basis of that judgment is too narrow and is outweighed by the larger rational considerations producing the incontinent action and apparently rendering it rational.” (Audi 1990, 277)

Für Nomy Arpaly sind willensschwache Akteure zwar notwendigerweise nicht gänzlich rational, da sie eine interne Inkohärenz aufweisen, dennoch können sie unter bestimmten Umständen rationaler agieren als Personen, die in denselben Situationen nach ihrem besten Urteil handeln.

„[...] I would like to argue that sometimes, an agent is more rational for acting against her best judgment than she would be if she acted in accordance with her best judgment. I still agree that every agent who acts against her best judgment is, as an agent, less than perfectly rational, as the schism between best judgment and desire indicates a failure of coherence in her mind. However, I will argue that it is not always the case that an agent is less rational (and less coherent) in following the desire than she is in following her best judgment. In fact, I will show, there are cases where following the best judgment would make the agent

---

<sup>477</sup> „[...] [A] rational action is one that is grounded, in the right way, in sufficiently good reasons of the agent, regardless of whether they figure in premises leading to the making of a judgment that favours that action.” (Audi 1990, 278f.)

<sup>478</sup> „Once we see that there is nothing sacrosanct about a practical judgment simply in virtue of its providing a directive whose violation is incontinent, we can also question whether the discord between such a judgment and action against it must make the action irrational. There is *something* amiss when incontinence occurs, but why must the action be the locus of our main criticism, or deserve the charge of irrationality? The action may, after all, yield the agent great pleasure – or, in less typical cases, a high degree of conformity with some cherished ideal – in a very efficient way and with no foreseeable bad effects. If it does, and if the judgment (however firmly held at the time) is itself one the agent would not accept on careful reflection, why must the action be, on balance, irrational? I cannot see that it must be.“ (Audi 1990, 275)

significantly irrational, while acting akratically would make her only trivially so – and as rational as most of us ever are.” (Arpaly 2000, 491)

Das Argument für die Möglichkeit der Rationalität willensschwacher Handlungen kann folgendermaßen verstanden werden:

- (P1) Wenn eine Person das tut, wozu sie tatsächlich am meisten Gründe hat, dann handelt sie rational.
- (P2) Es ist möglich, dass Personen glauben, dass sie am meisten Gründe haben, F zu tun, obwohl sie tatsächlich am meisten Gründe haben, eine alternative Handlung G zu tun.
- (P3) Es ist möglich, dass Person in Situationen wie (2) willensschwach handeln, also G tun, obwohl sie glauben, dass es besser wäre, F zu tun.
- (K) Also: Es ist möglich, dass Personen, auf die (2) und (3) zutreffen, rational handeln.

Der zentrale Punkt an dieser Argumentation ist, dass der Umstand, dass eine Person ein Urteil als besseres Urteil ansieht, das Urteil nicht wahr macht, d.h. dafür sorgt, dass es *tatsächlich* die beste mögliche Handlung der Person bezeichnet. Für die Wirklichkeit dieser Fälle wird unter anderem dadurch argumentiert, dass Emotionen als propositionale Zustände gedeutet werden, welche der Person, die sie hat, Gründe für ihr Handeln erschließen können. Emotionen erfüllen also auch eine epistemische Funktion, indem sie Werte entdecken, welche wiederum Gründe sein können für verschiedene Handlungen, diese Werte zu realisieren. Problematisch an dieser Argumentation ist, dass in diesen Fällen nicht klar ist, ob die Person tatsächlich *aufgrund von* Willensschwäche rational handelt. Es wäre, so ein Einwand, in diesen Situationen sicherlich rationaler, wenn die Person ihren Glauben darüber, was sie tun sollte, änderte, und dann entsprechend handelte, als wider ihr besseres Urteil zu handeln. Willensschwäche sei daher in solchen Situation nur im Vergleich rationaler als Willensstärke, aber nicht selbst rational. Wer also zeigen will, dass willensschwaches Handeln unter bestimmten Bedingungen rational sein kann, muss daher auch zeigen, dass in solchen Fällen Willensschwäche ein notwendiger Mechanismus ist, damit die Person rational handeln kann, und kein bloß optionales Element. Ein Beispiel: Manuel studiert Psychologie, weil er glaubt, dass er am besten Psychologie studieren soll, weil dies seinem Vater gefällt, obwohl in Wirklichkeit alles dafür spricht, dass er Mathematik studiert, weil er sehr begabt dafür wäre und es ihm Freude bereiten würde. Eines Abends steht er vor der Alternative, entweder für eine alles entscheidende Prüfung zu lernen oder einen Vortrag eines berühmten Mathematikers zu hören, den dieser an diesem Abend an seiner Universität hält. Manuel glaubt, dass er alles in allem in die Bibliothek gehen sollte, um für die Prüfung zu lernen. Auf dem Weg trifft er Lara, die zu dem Vortrag des Mathematikers geht. Entgegen seinem besseren Urteil entscheidet sich Manuel, mit Lara zu dem Vortrag zu gehen anstatt zu lernen. Aufgrund des

Vortrages sieht er ein, dass seine wahre Berufung in der Mathematik statt in der Psychologie liegt, und wechselt dementsprechend sein Studium.

Drei Möglichkeiten stehen Manuel offen: 1) er ist willensstark, folgt seinem falschem besseren Urteil und lernt für die Prüfung; 2) er ändert sein Urteil und glaubt nun, dass er Mathematik statt Psychologie studieren sollte; 3) er handelt willensschwach, indem er wider sein besseres Urteil zu dem Vortrag geht anstatt für die Prüfung zu lernen. Man könnte meinen, dass hier Willensschwäche zwar rationaler ist als Option 1), doch alles in allem der rationalste Weg Option 2) wäre. Dies aber nur dann, wenn Manuel sein Urteil von sich aus, ohne Hilfe von Willensschwäche, ändern könnte. In diesem Beispiel jedoch hilft Manuels Willensschwäche ihm, sein falsches Urteil zu ändern, und ist damit Teil eines Mechanismus, der es unvollkommen rationalen Wesen ermöglicht, irrational gebildete Urteile zu entlarven und deren Effekte zu mindern. Uwe Seel hat dies in seinem Lob der Willensschwäche folgendermaßen formuliert:

„Denn jede der starken Festlegungen, denen wir vorbehaltlos folgen, könnte gleichwohl schwach sein oder sich als schwach erweisen: sie könnte, entgegen unserer gegenwärtigen Meinung, nicht darauf zielen, was für uns das Beste ist. Schließlich steht auch das, was wir zu einer Zeit vorbehaltlos wollen, unter dem Vorbehalt, dass das hierbei Erstrebt für uns auch weiterhin unbedingt lohnend ist. Könnte sich dies nicht ändern, wären wir Gefangene unseres Wollens; könnten wir zu dieser Änderung nicht vorsätzlich beitragen, wäre es kein freies Wollen, durch das wir uns jetzt gebunden wissen. Die grundsätzliche Instabilität unserer praktischen Selbstverständnisse und Selbstverhältnisse ist nicht nur ein Mangel unserer Orientierungen; sie ist auch eine Offenheit für Orientierungen, die wir noch nicht gewonnen haben; sie ist darüber hinaus eine Freiheit der Orientierung, die sich nicht im Haben bestimmter Orientierungen erfüllt. Kein Zweifel, Willensschwäche ist aus der Perspektive der Handelnden zunächst einmal in jedem Fall eine Schwäche. Aber sie ist nicht nur das; sie ist manchmal auch ein erster Schritt in eine über kurz oder lang als richtig erkannte Richtung. Außerdem ist sie ein genereller Indikator der Freiheit, die wir auch gegenüber unseren vermeintlich oder tatsächlich besten Absichten haben.“ (Seel 2002, 241f.)

Willensschwäche kann demnach als eine Art Sicherheitsmechanismus funktionieren, um epistemische Unvollkommenheit gegenüber evaluativen Verhältnissen auszugleichen. In solchen Fällen hebt die eine Unvollkommenheit (Willensschwäche) die Wirkungen der anderen (Fehlurteil) auf.

## **VI.2 Willensschwäche als relatives Laster**

Da die normative Analyse von willensschwachen Einzelhandlungen zu keinem eindeutigen Ergebnis führen kann, da sie vorwiegend von den Umständen und Einzelheiten der jeweiligen Handlung abhängen als von der Struktur der Willensschwäche, ist es besser, Willensschwäche als Charakterzug, als Phänomen mit gewisser Beständigkeit, zu untersuchen. Nach Thomas Hill sollte ein moralphilosophische Untersuchung von Willensschwäche als Charakterzug klären, um welchen Fehler es sich bei Willensschwäche handelt, weshalb sie vermieden werden sollte und in

wie fern sie überhaupt moralisch relevant ist.<sup>479</sup> Methodisch geht er dabei umgekehrt zur üblichen handlungstheoretischen Analyse von Willensschwäche vor: Am Anfang der Untersuchung steht die alltägliche moralische Bewertung des willensschwachen Charakters, an diese anknüpfend folgt die Unterscheidung verschiedener Formen solcher Charaktere und schließlich die moralphilosophische Rechtfertigung und Erläuterung der moralischen Urteile. Die handlungstheoretischen Voraussetzungen sind minimal: wichtig ist vor allem die Möglichkeit, dass sich das beste Urteil und der Wille einer Person nicht notwendig decken müssen.

Willensschwäche ist nach Hill zu unterscheiden von angrenzenden Phänomenen der 1) Willenlosigkeit: die Person verfügt nicht über das Bündel an Fähigkeiten, welchen den Willen ausmachen;<sup>480</sup> 2) Handeln wider das nach eigenem Urteil Beste: Personen können willensschwach sein, auch wenn sie ihrem besten Urteil folgen, wenn die Revision des besten Urteils irrational ist;<sup>481</sup> und 3) Mangelnde Willenskraft: die Person verfügt zwar prinzipiell über einen Willen, also einem Bündel an Fähigkeiten zur Kontrolle und Bestimmung des eigenen Handelns, doch kann die Person zeitweise ihren Willen nicht adäquat einsetzen.<sup>482</sup> Zentral für Willensschwäche ist für dieses Konzept ein Verständnis dessen, was es heißt, einen Willen zu haben. Ich verstehe dabei Hills Ausführungen als eine vorwiegend funktionale Charakterisierung des Willens, also als Erläuterung der wesentlichen Funktionen und Fähigkeiten, über die jemand verfügt, der einen Willen hat.<sup>483</sup> Dies sind: 1) Kenntnis grundlegender kausaler Zusammenhänge; 2) Kenntnis der Möglichkeit des kausalen Einflusses eigener Handlungen auf Zustände in der Welt; 3) Wünsche bzw. Pro-Einstellungen gegenüber spezifischen Zuständen; 4) Kenntnis möglicher innerer und äußerer Hindernisse und Konflikte; 5) Persistenz und Konstanz des Willens in Relation zu diesen Hindernissen; 6) Planung und Kontrolle komplexer Handlungen und die Beeinflussung zukünftiger Überlegungen und Entscheidungen im Lichte vergangener Entscheidungen. Willensschwäche erläutert Hill nun folgendermaßen:

„Willensschwache sind zwar nicht willenlos oder ohne Willenskraft, aber sie sind Personen, die wiederholt und häufiger als normal (a) unzulängliche Anstrengungen unternehmen, (b) mit zu geringer Entschiedenheit Entschlüsse fassen, (c) ihre Vorsätze – sei’s mit oder ohne vorheriges Ringen – brechen und/oder (d) allzu leicht ihre Vorhaben aufgeben.“ (Hill 2005, 182)

Diese Erläuterung entspricht im Großen und Ganzen einem Verständnis von Willensschwäche als exekutive Willensschwäche, erweitert um die Dimension der Persistenz und Häufigkeit entsprechender willensschwacher Handlungen, um eine erfolgreiche Zuschreibung eines willensschwachen Charakters zu ermöglichen. Hill ist sich dabei bewusst, dass wesentliche Begriffe

---

<sup>479</sup> Vgl. Hill 2005, 168.

<sup>480</sup> Vgl. Hill 2005, 173–176.

<sup>481</sup> Vgl. Hill 2005, 176–179.

<sup>482</sup> Vgl. Hill 2005, 179–182.

<sup>483</sup> Vgl. Hill 2005, 174ff.

dieser Erläuterung von Willensschwäche vage und unscharf sind und einen dynamischen Aspekt aufweisen. So hängen die Standards für ungenügende Anstrengungen und dem Fassen zu schwacher Vorsätze von dem jeweiligen Handlungsbereich ab. Klar ist, dass in diesem Modell die zeitliche Dimension einer Person eine hervorragende Rolle einnimmt: nur vor dem Hintergrund eines halbwegs stabilen Selbstbildes bzw. einer Auffassung von sich als Person mit stabilen Eigenschaften ist die Zuschreibung von Willensschwäche überhaupt erst sinnvoll. Weshalb ist Willensschwäche nun ein Charakterzug, der vermieden werden sollte? Hills Antworten auf diese Frage lassen sich in prudentielle und moralische Gründe gegen Willensschwäche einteilen:

1) Prudentielle Gründe: Willensschwache Personen erleiden unter normalen Umständen verschiedene Arten von Nachteilen aufgrund ihrer Willensschwäche.<sup>484</sup> Es sei unwahrscheinlicher, dass sie ihre eigenen Präferenzen erfüllen, zudem seien sie enttäuscht über Misserfolge und empfinden oft Reue wegen vertaner Chancen. Auch sehen sie sich wegen ihrer Willensschwäche negativen Bewertungen von ihrem Umfeld ausgesetzt. Diese Gründe sprechen jedoch nur kontingenterweise gegen Willensschwäche, da es sein kann, dass eine Person noch mehr Nachteile erleiden oder gravierendere Fehler begehen würde, wenn sie willensstark wäre, auch wenn dies unter normalen Umständen wohl eher zweifelhaft ist. Daneben ist Willensschwäche dazu geeignet, die Bedingungen für Selbstachtung und den Selbstrespekt einer Person für sich zu untergraben.<sup>485</sup> Auf diesen Punkt werde ich noch später genauer eingehen.<sup>486</sup>

2) Moralische Gründe: Abgesehen von der moralischen Relevanz von Selbstachtung und Selbstrespekt an sich könne Willensschwäche sowohl zur Verletzung von Pflichten gegenüber anderen als auch solchen gegenüber sich selbst beitragen.<sup>487</sup> Willensschwache Personen versagen öfters, diesen Pflichten im erforderlichen Ausmaß nachzukommen, da diese oft mit entsprechenden Anstrengungen und einer gewissen Konstanz im Handeln verbunden sind. Willensschwäche mache es außerdem unwahrscheinlich, supererogatorische Leistungen zu vollbringen, also Handlungen zu setzen, die über das moralisch unbedingt erforderliche Maß hinausgehen. Auch im moralischen Bereich liegt die letztliche Bewertung der Willensschwäche einer Person jedoch an der Qualität der übrigen Tugenden, welche die Person aufweist. Willensschwäche ist damit nur ein relatives bzw. bedingtes Laster und Willensstärke nur eine relative bzw. bedingte Tugend.

„Alle diese Argumente (außer vielleicht dem allerletzten) [Argumente, die Willensschwäche als Verstoß gegen moralische Verpflichtungen gegenüber sich selbst ausweisen] deuten den Charakterzug der Willensschwäche nicht als an und für sich böse oder lasterhaft, sondern als einen Zug, der unter normalen Umständen beharrlich dazu neigt, eine den Anforderungen und Idealen der Moral entsprechende Lebensführung zu stören. Das mag zwar genügen, um die Willensschwäche zu den moralischen „Lastern“

---

<sup>484</sup> Vgl. Hill 2005, 186f.

<sup>485</sup> Vgl. Hill 2005, 187.

<sup>486</sup> Vgl. Kap. VI.4.

<sup>487</sup> Vgl. Hill 2005, 188–190.

zu rechnen, aber es scheint für sich genommen nicht auszureichen, um sie mit Recht in der gleichen Weise zu tadeln wie Grausamkeit, Ungerechtigkeit und Unredlichkeit.“ (Hill 2005, 190)

Zu einem ähnlichen Ergebnis der tugendethischen Bewertung von Willensschwäche gelangt Robert C. Roberts. Er unterscheidet zwischen motivationalen bzw. substantiellen Tugenden (motivational bzw. substantive virtues) und Tugenden der Willenskraft (virtues of will power).<sup>488</sup> Zu den ersten gehören Tugenden wie Mitgefühl, Ehrlichkeit, Gerechtigkeit u.a., die sich jeweils auf spezifische Handlungsmotive beziehen. Zu den zweiten gehören Tugenden wie Mut, Geduld und Selbstkontrolle und sie beziehen sich nicht auf Motive zu Handlungen, sondern auf Fähigkeiten, Handlungen erfolgreich durchzuführen. Nur in untypischen Fällen handelt jemand, *um* geduldig zu sein oder um der Geduld willen, sondern normalerweise ist jemand geduldig, um ein bestimmtes Ziel zu meistern. Diese Tugenden sind zu einem wichtigen Teil Fähigkeiten (skills) des Selbstmanagements und nur bei Wesen nötig, die prinzipiell mit einander widerstreitenden Handlungsmotiven und plötzlich auftretenden Handlungsimpulsen konfrontiert sind.<sup>489</sup> Geduld ist deswegen nötig, da Personen während der Durchführung einer Handlung mit plötzlich auftretenden widerstreitenden Handlungsimpulsen konfrontiert sein können, die sie von der gerade durchgeführten Handlung ablenken würden. Mut ist deswegen nötig, da bestimmte Emotionen in Situation auftreten können, welche zur Flucht oder zur Aufgabe der Handlung drängen würden.<sup>490</sup> Da diese Tugenden Fähigkeiten sind, sind sie zu einem großen Teil lehr- und lernbar. Auf basaler Ebene werden diese Fähigkeiten habituell verinnerlicht und kaum strategisch bewusst eingesetzt.

Entscheidend für die tugendethische Bewertung ist ihr bedingter Charakter: nur im Zusammenspiel mit substantiellen bzw. motivationalen Tugenden sind sie ethisch wertvoll. Der geduldige Einbrecher und der furchtlose Auftragskiller sind nur extreme Beispiele dafür, dass diese Tugenden, als Fähigkeiten betrachtet, an sich ethisch neutral sind. Dennoch stehen sie auf zumindest zwei Weisen in noch engerer Verbindung zu motivationalen Tugenden: Erstens sind typischerweise Personen mit ausgeprägtem Willen auch Personen, die sich besonders um etwas sorgen. Die Motivation, geduldig zu sein, ist schließlich verknüpft mit der Motivation, etwas durch Geduld zu erreichen, und nicht um der Geduld selbst willen. Zweitens verfügen willensstarke Personen auch typischerweise über Selbstvertrauen:

---

<sup>488</sup> Vgl. Roberts 1984, 228f.

<sup>489</sup> Vgl. Roberts 1984, 238ff.

<sup>490</sup> „Any of the virtues of will power may involve power over emotions. Self-control can be, among other things, a mastery of anger, resentment, malicious joy, scorn, hatred and envy. Where patience is directed to a task, the emotion with which it typically has to do is boredom. (A related virtue is tenaciousness or perseverance, whose emotional adversary seems to be primarily discouragement or some degree of hopelessness.) Where patience is used in dealing with people, emotions typically mastered are anger and resentment and boredom. The emotions that courage relates to are fear and anxiety.“ (Roberts 1984, 245)

„There is a correlation between people's ability to handle their adverse inclinations and the confidence of their belief in their ability to determine their own destiny. Of course it is typical of skilled people in general that they are confident of themselves in respect of the skill in question. And so, since self-mastery is so fundamental to the successful conduct of a human life, it might seem that self-confidence is a by-product of the virtues of will power rather than an element of them. But while self-mastering skills clearly cause self-confidence, it is equally true that self-confidence grounds the exercise of self-mastery skills.” (Roberts 1984, 247)

Beide Analysen deuten Willensschwäche damit als relatives Laster, das unter normalen Umständen sowohl aus Klugheitsgründen als auch aus moralischen Gründen zu unterlassen ist.

### VI.3 Willensschwäche als fehlendes Selbstvertrauen

Edward Hinchman verfolgt in *Receptivity and the will* (2009) das Ziel, eine Erläuterung von motivationaler und exekutiver Willensschwäche innerhalb des theoretischen Rahmens eines motivationalen Internalismus zu geben. Unter dem Willen versteht er eine rezeptive Kompetenz, sich von früheren Urteilen oder Entscheidungen im gegenwärtigen Handeln leiten zu lassen. Daher ist Willensschwäche eine Form von mangelhafter Selbstsorge in Form von mangelndem Selbstvertrauen gegenüber eigenen evaluativen Urteilen oder eigenen getroffenen Entscheidungen (Intentionen). Nach Hinchman taucht Willensschwäche als Bruch zwischen Urteil und Intention (motivationale Willensschwäche) sowie zwischen Intention und daran anschließendem Handeln (exekutive Willensschwäche) auf.<sup>491</sup> Motivationale Willensschwäche widerspreche zwei Formen von Internalismus, einem Urteils-Internalismus (judgment internalism) und einem Willens-Internalismus (volitional internalism): Ersterer behauptet eine notwendige Verbindung zwischen bestimmten evaluativen Urteilen und Intentionen (forward connection), zweiterer zwischen Intentionen und bestimmten evaluativen Urteilen (backward connection). Exekutive Willensschwäche widerspreche einem Ausführungs-Internalismus (resolve internalism), der eine notwendige Verbindung zwischen Intentionen und deren Ausführung in einer Handlung behauptet.<sup>492</sup>

Der von Hinchman vertretene Internalismus ist konditional, d.h. er formuliert eine zusätzliche Bedingung, die gegeben sein muss, damit auf das evaluative Urteil tatsächlich eine Intention folgt bzw. damit auf eine Intention tatsächlich eine Handlung folgt. Die Notwendigkeit oder der nicht-empirische Charakter der besagten Verbindungen ist dabei für Hinchman Kennzeichen internalistischer Theorien.<sup>493</sup> Dieses Verständnis von Internalismus ist jedoch zu weit, da inter-

---

<sup>491</sup> Die erste Form bezeichnet Hinchman als *incontinence*, die zweite als *weakness*. Vgl. Hinchman 2009, 395f. Ich bleibe hier bei den zu Anfang eingeführten Bezeichnungen motivationale und exekutive Willensschwäche.

<sup>492</sup> Hinchman merkt an, dass theoretisch auch eine vierte Form von Internalismus möglich sei, die eine notwendige Verbindung zwischen (intentionalen) Handlungen und Intentionen behauptet. Diese Position werde jedoch nicht unter dem Namen Internalismus diskutiert. Vgl. Hinchman 2009, 423f. fn. 3.

<sup>493</sup> „To call these relations ‘internal’ is to say that they’re in some sense necessary or non-empirical.” (Hinchman 2009, 396)

nalistische Theorien zusätzlich zeigen sollten, dass die notwendige Verbindung den evaluativen Urteilen intern ist, d.h. an diesen selbst liegt.<sup>494</sup> In diesem Sinne ist wohl die Abgrenzung von Hinchman zu externalistischen Theorien zu verstehen:

„The threat in each case is that we cannot understand how such counternormative agency is possible without positing a mediating faculty of the will with one or both of these features: to accommodate incontinence, it must function independently from judgment; to accommodate weakness, it must be exercised with greater or lesser degrees of ‘strength.’ Since such a faculty of the will would falsify the relevant internalist thesis, any view that posits such a faculty will count as a species of externalism.” (Hinchman 2009, 396)

Eine Möglichkeit, Willensschwäche zu erläutern, ist demnach, ein von evaluativen und praktischen Überlegungen unabhängiges Vermögen, einen Willen, einzuführen, dessen Funktion gerade in dem Transfer evaluativer Urteile zu motivational wirksamen Intentionen sowie dem Aufrechterhalten derselben liegt. Problematisch an dieser Abgrenzung ist, dass die Theorie von Hinchman diesen Ansprüchen nicht genügt. Grundlegend an seinem Konzept ist die Annahme eines nicht-deliberativen Vermögens, das sensitiv ist für Evidenzen, die Misstrauen rechtfertigen, und im gegebenen Fall eine bereits abgeschlossene Überlegung wieder eröffnen oder die Ausführung einer Intention aussetzen kann. Da das Vermögen nicht-deliberativ ist, und Theorien, die ein solches nicht-deliberatives Vermögen voraussetzen, externalistisch sind, sollte seine Theorie seinem eigenen Verständnis nach als externalistisch und nicht als internalistisch verstanden werden.

„What’s not easy, on my account, is trusting reasonably in the face of the twin phenomena of unruly desires and less than perfectly trustworthy judgment or intention. The difficulty lies in reopening deliberation *when but only when* the untrustworthiness of the judging or intending self requires it. Since the question is whether to redeliberate, it must be settled through the exercise of a non-deliberative form of intelligence. I hold that this intelligence is most fundamentally a counterfactual sensitivity to evidence of untrustworthiness in the judging or intending self: if you had evidence that your judgment or intention had been untrustworthily formed you would not simply have followed through on it.” (Hinchman 2009, 397)

Die willensschwache Person vertraut faktisch nicht ihrem besten Urteil oder ihrer auf einem solchen beruhenden Intention, und formt eine Intention bzw. handelt, anstatt dieses Misstrauen zum Anlass einer erneuten Überlegung über die Gründe ihres Handelns zu nehmen. Nur Intentionen, die tatsächlich durch die eigenen guten Gründe informiert sind, entsprechen dem eigenen Willen, wie auch nur solche Handlungen, die auf solchen Intentionen beruhen.<sup>495</sup> Das Selbstvertrauen in die eigenen Urteile und Intentionen soll dabei sowohl die rationale Rechtfertigung als auch die motivationale Effizienz der eigenen Urteile auf das Handeln

---

<sup>494</sup> Vgl. Kap. III.5.

<sup>495</sup> „My thesis, then, is twofold: your choice or intention functions as your will when it expresses an executive authority that derives from its trusting receptivity the verdict of your judgment, and your behavior manifests your will when it manifests a trusting receptivity to the influence of your choice or intention. Choice or intention registers thus receptively the authority of trustworthy judgment, making it motivationally efficacious. My main argument is that this thesis best explains the possibility of the two species of akratic break.” (Hinchman 2009, 405)

übertragen. Motivational effektiv sind Urteile nach diesem Modell deshalb, da ihnen vertraut wird, und rationale Autorität haben sie inne, da ohne ein gewisses Maß an Selbstvertrauen praktische Überlegungen niemals zu einem Abschluss gelangen könnten. Der Wille als rezeptives Vermögen greift dabei nur dann ein, wenn Evidenzen vorliegen, dass dem eigenen Urteil bzw. der eigenen Intention nicht mehr vertraut werden sollte.

Die Erläuterung von exekutiver Willensschwäche steht im direkten Gegensatz zu derjenigen von Richard Holton. Bei Hinchman liegt der Fehler der willensschwachen Person darin, trotz des fehlenden Selbstvertrauens nicht noch einmal das eigene Handeln zu überdenken, bei Holton liegt er darin, dass willensschwache Personen über Intentionen wieder nachdenken, anstatt sich auf diese zu verlassen und ihnen einfach zu folgen.<sup>496</sup> Dieser Widerspruch lässt sich auflösen, wenn man annimmt, dass der charakteristische Fehler von willensschwachen Akteuren verschieden ist je nach Form der gezeigten Willensschwäche: Das Problem des exekutiv Willensschwachen ist, dass er zu oft in Fällen überlegt, in denen er lieber handeln sollte. Das Problem des motivational Willensschwachen ist dagegen, dass er zu handeln bereit ist, ohne vorher noch einmal zu überlegen. Die gemeinsame Bezeichnung beider Fehler als mangelndes Selbstvertrauen bei Hinchman verdeckt diese Asymmetrie.

Hinchmans weiteres Modell basiert zudem auf einer problematischen Reduktion der zentralen Begriffe ‚Entscheidung‘, ‚Urteil‘ und ‚Intention‘ auf Dispositionsbegriffe.<sup>497</sup> Eine Entscheidung zu fällen sei dabei das Haben einer komplexen Disposition, sich so zu verhalten, als hätte man ein bestimmtes Urteil getroffen und wäre dabei praktisch weise verfahren.

„(C) S chooses to  $\varphi$  at t iff S accepts<sup>498</sup>: that she judges (all things considered) that she ought to  $\varphi$  at t and in so judging is practically wise.“ (Hinchman 2009, 412)

---

<sup>496</sup> Vgl. Holton 2009: „The basic idea is this. Just as it takes skill to form and employ beliefs, so it takes skill to form and employ intentions. In particular, if intentions are to fulfil their function, they need to be relatively resistant to reconsideration. Weakness of will arises, I shall suggest, when agents are too ready to reconsider their intentions.“ (Holton 1999, 241)

<sup>497</sup> „My account of practical judgment will be reductive: it will reduce judging that you ought to  $\varphi$  to a complex of dispositions, none of which on its own possesses the rational authority characteristic of practical judgment. The practical judgment that you ought to  $\varphi$  neither is nor entails, of course, the simple disposition to  $\varphi$  – since that would rule out the akratic failure to be disposed to act as you judge best. Nor is it the mere second-order disposition to treat yourself as disposed to  $\varphi$  – since that would omit the rational dimension of the self-relation. It does, however, build on that second-order disposition. Practical judgment adds to the second-order disposition a further disposition to treat the fact that you have that second-order disposition as giving you a reason to realize it in first-order action. And not just any reason: you’re disposed to treat the second-order disposition as giving you a preemptive reason to act, which I define as a reason to act without further deliberation. [...] In sum: a practical judgment is the disposition to treat the fact that you are disposed to treat yourself as disposed to  $\varphi$  as a preemptive reason to  $\varphi$ .“ (Hinchman 2009, 413)

<sup>498</sup> ‚Etwas zu akzeptieren‘ wird mit ‚so zu tun, als wäre es der Fall‘ als Handlungsdisposition analysiert. „Call being disposed to act as if p ‘accepting’ that p.“ (Hinchman 2009, 412)

Ein Urteil zu fällen, dass man eine Handlung alles in allem tun sollte, wird reduziert auf eine komplexe Disposition, sich so zu verhalten, als ob man einen besonderen Grund (preemptive reason)<sup>499</sup> hätte, diese Handlung zu tun, *weil* man sich so verhält, als hätte man tatsächlich solch einen Grund.

„(PJ) S judges (all things considered) that she ought to  $\varphi$  at t iff S accepts: (i) that she has a preemptive reason to  $\varphi$  at t, (ii) that she accepts (i), and (iii) that (i) because (ii).” (Hinchman 2009, 413)

Schließlich wird die Intention, etwas zu tun, verstanden als Erwartung und Wunsch eines früheren Selbst an sich, zu einem bestimmten späteren Zeitpunkt einen besonderen Grund (preemptive reason) zu haben, die Handlung aufgrund der Erinnerung an diesen Wunsch zu tun.

„(I)  $S_{F/A}$  intends to  $\varphi$  at t iff F expects and desires that A will at t have a preemptive reason to  $\varphi$  at that time simply through A’s memory of that desire.” (Hinchman 2009, 415)

Hinchman will mittels dieser reduktiven Analyse einerseits dem Einwand entgehen, sein Modell baue auf einem Bild von Handlungen auf, das den Handlungsprozess als strikte Abfolge zeitlich distinkter Phasen vom Urteil zur Entscheidung bzw. Intention und schließlich zur Handlung verstehe.<sup>500</sup> Innerhalb der dispositionalen Analyse macht eine solche Aufeinanderfolge keinen Sinn. Andererseits will er das spezifische normative Moment eines Urteils, dass man eine Handlung alles in allem gesehen tun sollte, und einer Intention, etwas zu tun, herausarbeiten. Etwas unterscheidet bloße Gedankenspiele von echten Urteilen, und bloße Wünsche von tatsächlichen Intentionen. Ein echtes Urteil darüber, was man tun soll, unterscheidet sich nach Hinchman von einem bloßen Gedankenspiel dadurch, dass sich die Person in diesem Urteil als Autorität akzeptiert, ihr eigenes Handeln zu bestimmen, d.h. sich selbst einen bestimmten Grund gibt, etwas in Zukunft *ohne weitere Überlegung* zu tun. Ob sie sich tatsächlich als solch eine Autorität akzeptiert, hängt jedoch davon ab, ob die Person auch tatsächlich so handelt, wie es ihrem Urteil entspricht – dies öffnet das Konzept für die Möglichkeit von Willensschwäche. Eine Intention, etwas zu tun, unterscheidet sich von dem bloßen Wunsch, etwas zu tun, dadurch, dass die Person zu dem Zeitpunkt des Handelns keine weitere Überlegung anstellen muss, was sie tun soll, sondern einfach die Intention ausführt, an die sie sich erinnert. Um sich hier als normative Autorität tatsächlich zu respektieren bzw. sich selbst zu vertrauen muss die Person auch tatsächlich tun, woran sie sich erinnert – auch hier verbleibt also eine Möglichkeit für exekutive Willensschwäche, nämlich nicht auszuführen, was man sich selbst vorgenommen hat. Die normative Autorität praktischer Rationalität wird nach dem Modell von Hinchman als

---

<sup>499</sup> Unter einer ‚preemptive reason‘ versteht Hinchman einen Grund dafür, etwas ohne weitere Überlegung zu tun. Vgl. Hinchman 2009, 413.

<sup>500</sup> Vgl. Hinchman 2009, 410ff.

Normativität des Selbstverhältnisses gedeutet – als gerechtfertigtes Selbstvertrauen und Selbstmisstrauen bzw. deren ungerechtfertigte Kehrseiten.

#### **VI.4 Willensschwäche und Selbstachtung**

Sowohl die tugendethische Betrachtung als auch die Betrachtung von Willensschwäche als fehlendes Selbstvertrauen führen dazu, das normative und evaluative Moment von Willensschwäche als Charakterzug in seiner Beziehung zu Selbstachtung zu analysieren. Innerhalb der Konzeption von Gerechtigkeit als Fairness von John Rawls kommt dem primären Gut der Selbstachtung eine zentrale Stellung zu.<sup>501</sup> Selbstachtung umfasst dabei 1) *Selbstwertgefühl*, d.i. den Glauben daran, dass die eigenen Vorstellungen des Guten auch tatsächlich gut sind, und 2) *Selbstvertrauen* in die eigene Handlungsfähigkeit, also die Fähigkeit, die eigene Vorstellung eines guten Lebens auch in die Tat umsetzen zu können. Fehlendes Selbstwertgefühl und mangelndes Selbstvertrauen untergraben zentrale Faktoren für ein gelungenes Leben, daher kommt Selbstachtung als Wert eine überragende Stellung im Leben einer jeden Person zu.

„When we feel that our plans are of little value, we cannot pursue them with pleasure or take delight in their execution. Nor plagued by failure and self-doubt can we continue in our endeavors. It is clear then why self-respect is a primary good. Without it nothing may seem worth doing, or if some things have value for us, we lack the will to strive for them. All desire and activity becomes empty and vain, and we sink into apathy and cynicism.” (Rawls 1971, ch. 67, 386)

Aus dieser Perspektive ist die schädliche Wirkung von Willensschwäche für das Selbstvertrauen und damit für die Selbstachtung einer Person ersichtlich. Willensschwache Akteure können ihre Konzeption des guten Lebens nur mangelhaft in die Tat umsetzen, da ihre Handlungen mit ihren evaluativen Urteilen über das aus ihrer Sicht Gute nicht übereinstimmen. Das normative Argument gegen Willensschwäche auf der Basis von Selbstachtung kann folgendermaßen formuliert werden:

- (P1) Selbstachtung ist ein hoher Wert im Leben einer jeden Person.
- (P2) Wenn etwas ein hoher Wert im Leben einer jeden Person ist, so sind prima facie alle Handlungen zu vermeiden, welche diesen Wert untergraben.
- (K1) Also: Wenn etwas die Selbstachtung einer Person untergräbt, dann ist es prima facie zu vermeiden.
- (P3) Willensschwäche untergräbt das Selbstvertrauen einer Person, d.h. die Fähigkeit, die eigenen Vorstellungen von einem guten Leben auch tatsächlich umzusetzen, und untergräbt damit einen Aspekt ihrer Selbstachtung.
- (K2) Also ist Willensschwäche prima facie zu vermeiden.

---

<sup>501</sup> Vgl. Rawls 1971, ch. 67.

Die Konklusion ist abgeschwächt auf die prima facie Vermeidung von Willensschwäche, da es (extreme) Umstände geben kann, in denen Willensschwäche zu einem gelungenerem Leben führt als Willensstärke. Dies sind Fälle, in denen die normativen und evaluativen Überzeugungen von Akteuren fundamental fehlerhaft sind, und so deren Ausführung zu katastrophalen Ergebnissen für das Leben der betreffenden Person führen würde. Auch hier zeigt sich der indirekte oder relative Charakter der Tugend der Willensstärke in Beziehung zu den tatsächlichen normativen und evaluativen Umständen einer Person im Unterschied zu den bloß vermeintlich bestehenden.

Trotz dieses generellen Argumentes gegen Willensschwäche als Charakterzug bleibt eine Bandbreite der Akzeptanz für willensschwaches Verhalten bestehen, geprägt vor allem in Beziehung zur jeweiligen Gesellschaft und den in ihr anerkannten sozialen Rollen: Das Ausmaß, in dem Willensschwäche die Selbstachtung einer Person untergräbt, hängt nämlich an den mit gesellschaftlichen Rollen verknüpften Standards, an denen die Person sich misst und an denen sie gemessen wird. Nach Luc Bovens bleiben einem durch akratische Konflikte geprägten Charakter zwei Möglichkeiten, seine Willensschwäche aktiv zu therapieren:<sup>502</sup>

1) *Steigerung der Verluste* aus willensschwachem Handeln: durch Maßnahmen der Selbstbindung können Personen ihr Umfeld oder sich selbst in einer Weise manipulieren, sodass willensschwache Handlungen mit herben Verlusten in verschiedenster Form verbunden sind, was wiederum die Motivation, willensschwache Handlungen zu unterlassen, steigert.<sup>503</sup>

„The serene side I know all too well: it is stacked with those who let no opportunity slide to raise the disutility of akratic acts. There are various ways of doing so. Some are masters at manipulating their surroundings: they construe their worlds such that it would be costly, time-consuming or just simply unpleasant to take up akratic acts. Others are masters at manipulating their characters: they come to think of themselves as selfdisciplined people in total disregard of who they really are. At least some do manage to pull themselves up by their bootstraps: akratic acts become all too costly, since they are threats to their self-image.” (Bovens 1999, 233)

2) *Steigerung der Gewinne* aus willensschwachem Handeln: anstatt willensschwache Handlungen mit zusätzlichen negativen Sanktionen auszustatten, können Personen auch soziale Rollen wählen, welche ein geringeres Maß an Selbstkontrolle für ein stabiles Maß an Selbstachtung voraussetzen.

„The ardent are most creative in finding ways to raise the utility of what would otherwise count as akratic acts. Their strategies are actually surprisingly similar to what is happening on the serene side. Some are masters at manipulating their environments. They are keenly aware what vices lure them into akrasia and they choose a way of life in which the exercise of these vices meets with rewards. [...] If you cannot or wish not to conquer your vices, you might as well organize your life in such a way that they turn into virtues. Others are masters at manipulating their characters. They come to construct an identity that places a disvalue on prudence and a value on spontaneity. [...] Consequently, their impulsive actions are no longer akratic, since each such action now has the added benefit that it strengthens their self-image.” (Bovens 1999, 234)

---

<sup>502</sup> Vgl. Bovens 1999.

<sup>503</sup> Maßnahmen der Selbstbindung diskutiert auch eingehend Jon Elster in Elster 2000.

Auch wenn durch solche Strategien das Maß an Willensstärke, das für ein ausreichendes Maß an Selbstachtung einer Person erforderlich ist, variiert, ist für die erfolgreiche Durchführung beider Strategien ein Minimum an Willensstärke erforderlich. Auch ein Rausch will organisiert sein, und eine Selbstbindungsmaßnahme muss tatsächlich eingegangen werden, damit sie greifen kann. Dies führt zur Frage, ob globale Willensschwäche überhaupt möglich ist: kann eine Person dermaßen desintegriert sein, dass sie niemals das intendiert, was sie als besser ansieht, oder niemals das tut, was sie sich eigentlich vorgenommen hat?

## VI.5 Ist globale Willensschwäche möglich?

Die Diskussion der normativen Signifikanz von Willensschwäche hat zu der Frage geführt, ob globale Willensschwäche überhaupt möglich ist und welche Gründe dagegen sprechen können.

„It is easy to feel the force of the intuition that global akrasia is impossible – but why? If global akrasia is impossible, it seems that must be because of some feature of the individual case.” (Hinchman 2009, 395)

Entgegen Hinchman möchte ich dafür argumentieren, dass die Unmöglichkeit von globaler Willensschwäche gerade nicht begründet wird an Eigenschaften von einzelnen Fällen von Willensschwäche, sondern anhand eines transzendentalen Argumentes anhand der Gesamtheit solcher Fälle. Unter *globaler motivationaler Willensschwäche* verstehe ich den Fall, dass eine Person evaluative oder normative bessere Urteile hält, aber *niemals* entsprechend dieser Urteile ihre Intentionen bildet. Unter *globaler exekutiver Willensschwäche* verstehe ich, dass eine Person niemals ihre Pläne in handlungsleitende Intentionen umsetzt, d.h. niemals so handelt, wie sie sich es zuvor vorgenommen hat. Das Argument baut darauf auf, dass eine Person nicht so interpretiert werden kann, als halte sie dauerhaft ihrem Verhalten widersprechende Urteile und Vorsätze. Positiv formuliert: Ein Teil ihres Verhaltens muss ihren Urteilen und Vorsätzen entsprechen, damit ihnen diese Urteile bzw. Vorsätze überhaupt zugeschrieben werden können. Das Argument ist inspiriert von Davidsons Argument, dass eine transendentale Voraussetzung für das Verstehen von Personen ist, davon auszugehen, dass ein Großteil ihrer Überzeugungen wahr ist.<sup>504</sup> Unser Verstehen des Verhaltens von anderen Personen ist an das Prinzip gebunden, das Verhalten als möglichst rational aus Sicht der handelnden Person zu interpretieren. Das Prinzip kann folgendermaßen formuliert werden:

Prinzip: Die Bedingung der Möglichkeit, das Verhalten von Akteuren zu interpretieren, ist, es als möglichst rational aus der Perspektive des Akteurs auszuweisen. Dies heißt grundsätzlich davon auszugehen, dass die Person das tut, was sie intendiert, und das intendiert zu tun, was sie für richtig oder der Situation angemessen hält. Zu jeder Handlung des Akteurs lässt sich daher eine Beschreibung seiner Intention konstruieren, welche die Handlung verständlich macht. Diese Beschreibung der Intention wiederum wird verständlich durch die Konstruktion eines Werturteils, das die entsprechende Handlung empfiehlt und dem

---

<sup>504</sup> Vgl. Davidson 1974. Es kann auch verstanden werden als transzendentales Argument für die Richtigkeit der ‚backward connection‘, vgl. Kap. IV.2.1.

die Intention entspricht. Das Werturteil rationalisiert demnach die Intention, welche wiederum die Handlung rationalisiert.

Eine Möglichkeit, für die Gültigkeit eines solchen Prinzips zu argumentieren, ist, Gedankenexperimente zu konstruieren, in welchem dieses Prinzip nicht gilt, und die entsprechenden Phänomene, also die interpretativen Leistungen, daher nicht aufweisen. Wie sähe eine Welt aus, in der das vorige Prinzip nicht gilt? Zuerst der Fall des konträren Gegenteils:

Konträr-Prinzip: Die Bedingung der Möglichkeit, das Verhalten von Akteuren zu interpretieren, ist, davon auszugehen, dass die Person *nicht* das tut, was sie intendiert. Weiters intendiert die Person *nicht* die Handlung, die sie selbst als richtig ansieht.

Da die Möglichkeiten, nicht das zu tun, was jemand intendiert, prinzipiell unendlich sind, und die Möglichkeiten, die Handlung zu intendieren, welche nicht durch das eigene Werturteil nahegelegt wird, ebenso unendlich ist, ist ein Verstehen von Handlungen in einer solchen Welt prinzipiell unmöglich. Wir wüssten stets nur, was die Person nicht zu tun intendierte, und was sie nicht selbst als zu intendierende Handlung empfiehlt, nicht, was sie eigentlich tun wollte und was sie glaubte, dass sie tun sollte.

Nun zum kontradiktorischen Gegenteil:

Kontra-Prinzip: Es ist nicht der Fall, dass wir grundsätzlich davon ausgehen können, dass Personen das tun, was sie zu tun intendieren, und auch nicht davon, dass sie die Handlung intendieren, die sie als zu tun durch ein Werturteil empfehlen. Dies heißt, dass wir grundsätzlich keine entsprechenden Intentionen und Werturteile für das Verständnis des Verhaltens anderer Akteure konstruieren sollten. Oder wir sollten dies durchschnittlich in der Hälfte der Fälle tun, in der anderen nicht. Mithin sollten wir zufällig entscheiden, ob wir das Verhalten als rational ausweisen oder nicht.

Auch diese Welt wäre radikal von unserer Welt verschieden. Das Verstehen fremden Verhaltens wäre partiell möglich, aber völlig unberechenbar. Dies wiederum hätte Auswirkungen auf die Möglichkeit geregelter sozialer Interaktion, welche unter Bedingungen des Kontra-Prinzips nicht möglich wären. Anhand dieses Prinzips wäre es nicht möglich, überhaupt konsistent Werturteile oder längerfristige Pläne zu konstruieren, von dem das Verhalten des Akteurs abweichen könnte. Diese Werturteile oder Pläne müssen sich schließlich in Verhalten äußern, z.B. in sprachlichen Äußerungen einer Person, damit sie ihnen zugeschrieben werden können. Diese Äußerungen im Verhalten sind aber selbst Gegenstand der Interpretation, und wenn diese geregelt ist durch das Kontra-Prinzip, dann sollten auch diese als bloß zufällig mit Werturteilen bzw. Plänen verbunden interpretiert werden. Auch unter dieser Voraussetzung verschwindet die Möglichkeit von Willensschwäche, da der Person keine konsistenten besseren Werturteile oder längerfristigen Pläne erfolgreich zugeschrieben werden könnten.

Anhand dieser Argumentation kann gezeigt werden, dass das ursprünglich transzendente Prinzip gültig ist, vorausgesetzt die tatsächlich stattfindenden Interpretationsleistungen von

Handlungen und Akteuren. Problematisch an dem ursprünglichen Prinzip ist, dass nicht klar ist, wie stark die Phrase ‚grundsätzlich‘ interpretiert werden soll. Im Folgenden kürze ich mit ‚rational‘ die mögliche Konstruktion einer Intention und eines dazu gehörenden Werturteils ab. Vier verschiedene Interpretationen sind möglich:

I1) Sehr starke Interpretation: Notwendigerweise und immer interpretieren wir Handlungen als rational, und sie sind es auch.

I2) Starke Interpretation: Notwendigerweise und immer interpretieren wir Handlungen als rational, aber sie sind es nicht immer.

I3) Schwache Interpretation: Notwendigerweise und grundsätzlich interpretieren wir Handlungen als rational, und sie sind es auch.

I4) Sehr schwache Interpretation: Notwendigerweise und grundsätzlich interpretieren wir Handlungen als rational, aber sie sind es nicht immer.

Nach I1 interpretieren wir nicht nur alle Handlungen immer als rational, sondern diese Handlungen sind auch stets rational. Dies schließt Willensschwäche nach gängigem Verständnis aus, da es jede Form von irrationalen Handeln ausschließt. Nach I2 gibt es zwar irrationale Handlungen, aber wir verstehen diese nicht als solche sondern rationalisieren sie. Diese Möglichkeit ist rein theoretisch, da wir nach dieser Variante niemals Fälle irrationalen Handelns feststellen könnten. Nach I3 interpretieren wir Handlungen nur grundsätzlich als rational, akzeptieren aber auch Fälle von irrationalen Handlungen. Diese kommen jedoch nicht vor. Auch diese Möglichkeit ist rein theoretisch. Variante I4 ist der Fall, dass wir grundsätzlich versuchen Handlungen als rational interpretieren – wir konstruieren eine Intention, die der Handlung entspricht, und konstruieren zudem ein Werturteil, dass die Intention empfiehlt –, falls diese Versuche jedoch scheitern, verstehen wir sie als minder rational bis irrational. Zudem kommen diese Handlungen nach I4 auch tatsächlich vor. Dieses Argument und das zugehörige transzendente Prinzip sind dabei unabhängig von externalistischen oder internalistischen Interpretationen von evaluativen und normativen Urteilen.<sup>505</sup>

## **Zusammenfassung**

Die Frage nach der Normativität von Willensschwäche ist die Frage nach den Normen und Werten, welche für oder gegen willensschwaches Handeln und Willensschwäche als Charakterzug sprechen. Die interne Irrationalität von Willensschwäche ist weithin akzeptiert und lässt sich interpretieren als Verstoß gegen ein rationales Prinzip der Bildung von Intentionen auf Basis von evaluativen Urteilen. Fälle von inverser Willensschwäche sprechen dafür, dass willensschwaches Handeln nicht unbedingt extern irrational ist. Willensschwäche als Charakterzug ist ein relatives Laster, da es unter normalen Umständen zu unklugem und unmoralischem Verhalten führt,

---

<sup>505</sup> Im Unterschied zur Argumentation von Sarah Stroud in Stroud 2003.

jedoch manchmal den Akteur vor groben Fehlern bewahren kann. Willensschwäche kann auch als Ausdruck von fehlendem Selbstvertrauen in die eigenen evaluativen Urteile und Überlegungen gedeutet werden. Dies führt zu einer Analyse von Willensschwäche als Untergrabung der eigenen Selbstachtung, einem hohen Wert im Leben eines jeden Akteurs. Da Selbstachtung für verschiedene Rollen und Lebensbereiche verschieden bedeutsam ist, ist der Unwert von Willensschwäche auch relativ zu diesen Rollen und Lebensbereichen. Eine transzendente Argumentation spricht schließlich dafür, dass Akteure zumindest über ein Minimum an Willensstärke verfügen müssen, da sie sonst nicht als Akteure mit eigenen Zielen und Werthaltungen interpretiert werden könnten. Letzteres spricht gegen die Möglichkeit globaler Willensschwäche ohne die Annahme eines deskriptiven motivationalen Internalismus.

## VII Willensschwäche als Kehrseite reflexiver Selbstbestimmung

In diesem Kapitel werde ich verschiedenen Themen der Arbeit nochmals aufgreifen und in eine Deutung von Willensschwäche als Kehrseite reflexiver Selbstbestimmung prägnant zusammenführen. Ich beginne mit dem Verhältnis von Evaluation und dem motivationalen Grundgerüst von Personen. Dann stelle ich das präferierte Modell praktischer Überlegung vor: Überlegungen *mit* praktischen Einstellungen anstatt Überlegungen *über* Praktisches. Dann untersuche ich die Rolle von Reflexion für die Selbstbestimmung von Personen. Zuletzt kläre ich die Frage der Bedeutung von Willensschwäche für ein gutes Leben.

### a) Evaluation und Motivation

Die willensschwache Person ist durch einen motivationalen Konflikt geprägt: dieser macht das Zentrum des Phänomens aus. Der Konflikt besteht einerseits aus einer Diskrepanz an evaluativer Einsicht und motivationaler Relevanz – die Person ist nicht proportional zu ihrem evaluativen Urteil motiviert, eine entsprechende Handlung zu setzen –, andererseits aus dem Konflikt zweier motivationaler Quellen, der Motivation für die dem besseren Urteil entsprechende Handlung und der Motivation für die willensschwache Handlung. Die Existenz dieses Konfliktes spricht für die Annahme disparater Quellen der Motivation, was wiederum gegen die Annahme einer Identität von Evaluation und Motivation spricht, da dies zu einer monistischen Theorie von Motivationsquellen führt. Sie spricht auch gegen Formen von deskriptiven Internalismen, welche eine strikte, da begrifflich notwendige Verbindung zwischen Evaluation und Motivation ansetzen. Dagegen setze ich das Modell eines deskriptiven Externalismus, der als hinreichende Quellen der Handlungsmotivation erst die Kombination von evaluativem Urteil mit einer entsprechenden Streben ansieht. Dies ist vereinbar mit einer *normativen* Interpretation des Verhältnisses von Evaluation und Motivation: die Person steht unter einer Anforderung der Rationalität, seine evaluativen Überzeugungen mit seinen motivationalen Zuständen in Einklang zu bringen. Dies kann die Person dadurch erreichen, indem sie korrekt überlegt, d.h. im Einklang mit den Regeln theoretischer und praktischer Rationalität.

### b) Praktische Überlegung

Die Frage ‚Was soll ich tun?‘ lässt zumindest zwei Antworten zu, die ich hier als *theoretische* und *praktische* Lösung bezeichnen möchte. Die theoretische Lösung beantwortet die Frage mit einem evaluativen oder normativen Urteil, welches *inhaltlich* eine mögliche Handlungsoption als besser oder als beste gegenüber anderen Optionen ausweist. Der mentale Akt gegenüber diesem praktischen Inhalt ist ein Glaubenszustand (mit gewisser Glaubensstärke). Aus motivational externalistischer Sicht ist dieser Glaubenszustand noch nicht mit einem motivationalen Zustand

identisch oder notwendig mit einem solchen verbunden. Die praktische Lösung beantwortet die Frage mit einer Intention bzw. Absicht, eine bestimmte Handlungsoption zu ergreifen und durchzuführen. Der mentale Akt ist hier kein kognitiver Zustand, sondern ein konativer. Als solcher ist er motivational wirksam hinsichtlich einer bestimmten Handlungsoption. In Abwesenheit von exekutiver Willensschwäche wird die Intention in Folge auch tatsächlich von der Person durchgesetzt. Der Begriff des ‚Entschlusses‘ ist mehrdeutig und schwankt zwischen der Bedeutung als Endpunkt einer theoretisch praktischen Überlegung, einem Urteil, und dem Endpunkt einer praktischen Überlegung, einer Intention. Diese Ambivalenz mag auch ein Grund sein für die recht weit auseinander gehenden Intuitionen gegenüber Beispielen von Willensschwäche.

Die zugehörigen Formen von Überlegungen sind a) rein theoretische Überlegungen: Überlegungen, welche kognitive Zustände mit anderen kognitiven Zuständen verbinden, welche in ihrem Inhalt nicht praktisch sind, also keine Handlungsoptionen betreffen und keinen evaluativen oder normativen Gehalt aufweisen; b) theoretisch praktische Überlegungen: Überlegungen über etwas Praktisches, welche rein kognitive Zustände mit praktischem Gehalt mit anderen kognitiven Zuständen verbinden; c) praktische Überlegungen: Überlegungen, welche konative Zustände mit kognitiven Zuständen oder anderen konativen Zuständen verbinden. Auf Basis dieser Unterscheidungen lässt sich verstehen, wie es möglich ist, dass eine Person dieselbe Frage ‚Was soll ich tun?‘ auf zwei voneinander abweichende oder sogar einander widersprechende Weisen beantworten kann. Eine willensschwache Person stellt eine theoretisch praktische Überlegung an und gelangt zu einem Urteil darüber, was sie in der gegebenen Situation am besten tun sollte. Ihre praktische Überlegung jedoch nimmt ihren Ausgangspunkt entweder nicht von dem Ergebnis ihrer theoretisch praktischen Überlegung, oder die willensschwache Person nimmt keine praktische Einstellung gegenüber diesem Ergebnis ein. Stattdessen deliberiert sie auf Basis eines Wunsches gegenüber einer alternativen Handlungsoption, einer Strebung, hin zu einer entsprechenden Handlungsintention für die willensschwache Handlung. Der willensschwachen Person fehlt eine praktische Prämisse, eine Intention gegenüber ihrem besseren Urteil, i.e. das zu tun, von dem sie glaubt, dass sie es tun sollte. Obwohl die beiden Arten von Überlegungen zu unterschiedlichen Ergebnissen gelangen können, sind diese doch miteinander verbunden. Die Verbindung ist dabei zweifach: einerseits *normativ*, andererseits *konstitutiv*. a) Normativ interpretiert ist es ein Gebot der Rationalität, sich im Gebrauch der Vernunft in seinen verschiedenen Formen nicht selbst zu widersprechen. Dieser normative Zusammenhang kann im Sinne von normativen Anforderungen (normative requirements) als rationale Anforderungen formuliert werden. Zentral ist dabei folgende Forderung: Es ist rational geboten, dass wenn ich glaube, dass ich F-en soll, ich auch intendiere zu F-en. Die vollständige

Formulierung umfasst noch einige spezielle Umstände und Probleme, die sich aus dem zeitlichen Aspekt ergeben. b) Konstitutiv ist der Zusammenhang zwischen theoretisch praktischen Überlegungen und praktischen Überlegungen für die Möglichkeit der Interpretation eines Wesens als handelnden Akteur: Eine global willensschwache Person, die zugleich zu theoretisch praktischen Schlüssen gelangt, diese jedoch niemals in die Tat umsetzt, kann nicht als handelnder Akteur aufgefasst werden. Dies hat sich in der Diskussion verschiedener möglicher transzendentaler Prinzipien für die Interpretation von Handlungen ergeben. Daraus folgt eine Unmöglichkeit von globaler Willensschwäche, da wir das Verhalten einer solchen Person nicht kohärent als von Intentionen und Werturteilen geleitet, mithin nicht als Handeln interpretieren könnten.

### c) Selbstbestimmung und Reflexion

Willensschwäche ist ein wesentlich reflexives Phänomen. Motivationale Willensschwäche entsteht aus der Diskrepanz zwischen einem evaluativen besseren Urteil und einer ihm widersprechenden Handlungsintention, exekutive Willensschwäche zwischen einem Handlungsplan und einer von ihm abweichenden aktuellen Handlungsintention. Exekutive Willensschwäche kann dabei auch von reflexiven Ressourcen des Akteurs profitieren, indem die Umsetzung eines Planes durch weitere, zuvor bereits als abgeschlossen behandelte Überlegungen hinausgeschoben wird. Selbstbestimmung ist die Bestimmung des eigenen Verhaltens durch die eigenen Wünsche, Evaluationen und darauf fußenden Intentionen des handelnden Wesens. Reflexiv ist eine Art der Selbstbestimmung, in der das Wesen zusätzlich zu ihrer Selbstbestimmung noch in einem Verhältnis zu ihrer Fähigkeit der Selbstbestimmung steht. Personen basieren ihr Verhalten nicht nur auf Wünschen, Evaluationen, Überzeugungen über die Welt und darauf basierende Intentionen, sondern auf Überzeugungen über eigene Wünsche und evaluative Überzeugungen. Wesen, die keine theoretisch praktischen Überlegungen anstellen, d.h. keine Reflexion darüber anstellen, welche Handlungsoptionen in gegebenen Situationen besser sind als andere, können keine motivationale Willensschwäche zeigen. Motivationale Willensschwäche ist die Kehrseite der Fähigkeit zur reflexiven Selbstbestimmung, ihr fehlerhaftes oder gescheitertes Gegenstück. Exekutive Willensschwäche dürfte auch in einem hohen Ausmaß mit der Fähigkeit zur reflexiven Selbstbestimmung einhergehen. Einerseits ist die Ausbildung von komplexen Handlungsplänen an entsprechende deliberative Kapazitäten des handelnden Wesens gebunden, andererseits können solche Deliberationsprozesse die erfolgreiche Ausführung von Handlungen erschweren. Der Sinn von Handlungsplänen ist schließlich neben der Koordination komplexer Tätigkeiten die Speicherung von entsprechenden Entscheidungen, die auf Basis von Überlegungen getroffen werden. Die Möglichkeit, solche Überlegungen jederzeit wieder aufzunehmen und damit den Handlungsfluss zu unterbrechen, kann nun dieser Speicherfunktion entgegenwirken und Formen

von exekutiver Willensschwäche erzeugen. In einigen Fällen von exekutiver Willensschwäche sind jedoch auch nicht-deliberative Faktoren ausschlaggebend, die demnach nicht mit der Fähigkeit reflexiver Selbstbestimmung unmittelbar verknüpft sind.

#### d) Willensschwäche und das gute Leben

Wenn Willensschwäche eine Kehrseite unserer Fähigkeit zur reflexiven Selbstbestimmung ist, stellt sich die Frage nach ihrem Beitrag für ein gelungenes oder gutes Leben. Generell ist Willensschwäche als Charakterzug zu meiden: die Diskrepanz zwischen der eigenen Vorstellung eines guten Lebens, das sich in evaluativen Urteilen ausdrückt, und den tatsächlichen Handlungen untergräbt die Selbstachtung der Person, welche wiederum fundamental ist für ein gelingendes Leben. Andererseits ist Willensschwäche auch Ausdruck von motivationalen Konflikten, in der eine Person sich befindet, und kann so die Person gewissermaßen über ihre eigenen Konflikte informieren. Außerdem zeigen Fälle von inverser Willensschwäche, dass über den Umweg einer willensschwachen Handlung die Person sich selbst davor bewahren kann, einem falschen evaluativen Urteil mit desaströsen Konsequenzen zu folgen. Willensschwäche fungiert in diesen Fällen als Sicherheitsmechanismus, um die epistemische Fehlbarkeit gegenüber evaluativen Sachverhalten auszugleichen.

## Bibliographie

Hinzufügungen in wortwörtlichen Zitaten sind durch Klammern ‚[ ]‘ kenntlich gemacht und stammen, wenn nicht anders vermerkt, von mir.

EE: Aristoteles (1962): Eudemische Ethik. Übers. v. Franz Dirlmeier. In: Aristoteles Werke in deutscher Übersetzung, Band 7. Hrsg. v. Ernst Grumach. Berlin: Akademie-Verlag.

NE: Aristoteles (2006): Nikomachische Ethik. Übersetzt und hrsg. v. Ursula Wolf. Reinbek bei Hamburg: Rowohlt-Taschenbuch-Verlag (rowohlt's enzyklopädie, 55651).

Ainslie, George

(2004): Breakdown of will. Cambridge: Cambridge Univ. Press.

(2005a): Die Delle in unserer Zukunftsbewertung. In: Spitzley 2005, 139–167.

(2005b): Précis of Breakdown of will. In: Behavioral and Brain Sciences 28 (5), 635–650.

Ameriks, Karl

(2006): Kant and motivational externalism. In: Klemme, Heiner F.; Kühn, Manfred; Schönecker, Dieter (2006) (Hg.): Moralische Motivation. Kant und die Alternativen. Hamburg: Meiner (Kant-Forschungen, 16), 3–22.

Ammereller, Erich

(2005): Die Gründe des Handelnden. In: Ammereller, Erich; Vossenkuhl, Wilhelm (2005) (Hg.): Rationale Motivation. Paderborn: Mentis, 57–83.

Anscombe, Gertrude E. M.

(1963): Intention. 2. ed. Cambridge: Harvard Univ. Press 2000.

(2011): Absicht. Aus dem Englischen von Joachim Schulte. Berlin: Suhrkamp (stw, 1978).

Aristoteles/Aristotle

(1985): Über die Bewegung der Lebewesen. Über die Fortbewegung der Lebewesen. Übers. und erläutert v. Jutta Kollesch. In: Werke in deutscher Übersetzung. Band 17. Zoologische Schriften II Teil 2/3. Hrsg. v. Hellmut Flashar. Berlin: Akademie-Verlag.

(1995): Über die Seele. Mit Einleitung, Übersetzung (nach W. Theiler) und Kommentar hrsg. v. Horst Seidl. Griechischer Text in der Edition von Wilhelm Biehl und Otto Apelt. Griechisch-Deutsch. Hamburg: Meiner (Philosophische Bibliothek, 476).

(2002): Nicomachean ethics. Rowe, Christopher; Broadie, Sarah (Hg.). Oxford: Oxford Univ. Press.

(2005): Eudemian ethics. Books I, II and VIII. Translated with a commentary by Michael Woods. 2. ed., reprint. Oxford: Oxford Univ. Press.

Arpaly, Nomy

(2000): On acting rationally against one's best judgment. In: Ethics 110 (3), 488–513.

Audi, Robert

(1990): Weakness of will and rational action. In: Australasian Journal of Philosophy 68 (3), 270–281.

Austin, John L.

(1956–57): A plea for excuses. In: Proceedings of the Aristotelian Society 57, 1–30.

Bach, Kent

(1978): A representational theory of action. In: Philosophical Studies 34 (4), 361–379.

(1980): Actions are not events. In: Mind 89 (353), 114–120.

(2010): Refraining, omitting, and negative acts. In: O'Connor & Sandis 2010, 50–57.

Baker, Lynne Rudder

(1995): Explaining attitudes. A practical approach to the mind. Cambridge: Cambridge Univ. Press.

Barnes, Jonathan

(1980): Aristotle and the methods of ethics. In: Revue Internationale de Philosophie 34, 490–511.

Bayertz, Kurt

(1995): Eine kurze Geschichte der Herkunft der Verantwortung. In: Bayertz Kurt (1995) (Hg.): Verantwortung. Prinzip oder Problem? Darmstadt: Wissenschaftliche Buchgesellschaft, 3–71.

Beckermann, Ansgar

(1977): Gründe und Ursachen. Zum vermeintlich grundsätzlichen Unterschied zwischen mentalen Handlungserklärungen und wissenschaftlich-kausalen Erklärungen. Kronberg: Scriptor (Monographien: Wissenschaftstheorie und Grundlagenforschung, 6).

(1985): Handeln und Handlungserklärungen. In: Beckermann, Ansgar (1985) (Hg.): Analytische Handlungstheorie. Band 2: Handlungserklärungen. Frankfurt am Main: Suhrkamp (stw, 489), 7–84.

Birnbacher, Dieter

(1995): Tun und Unterlassen. Stuttgart: Reclam (Universal-Bibliothek, 9392).

Bishop, John

(1989): Natural agency. An essay on the causal theory of action. Cambridge: Cambridge Univ. Press.

Bittner, Rüdiger

(2011): Aus Gründen handeln. Berlin: Walter de Gruyter.

- Björklund, Fredrik; Björnsson, Gunnar; Eriksson, John; Francén Olinder, Ragnar; Strandberg, Caj  
(2012): Recent work on motivational internalism. In: *Analysis* 72 (1), 124–137.
- Blackburn, Simon  
(1984): *Spreading the word. Groundings in the philosophy of language.* Reprint. Oxford: Clarendon Press 2004.
- Bostock, David  
(2000): *Aristotle's ethics.* Oxford: Oxford Univ. Press.
- Bovens, Luc  
(1999): The two faces of akratics anonymous. In: *Analysis* 59 (4), 230–236.
- Brand, Myles  
(1984): *Intending and acting. Toward a naturalized action theory.* Cambridge: MIT Press.
- Bratman, Michael E.  
(1987): *Intention, plans, and practical reason.* Cambridge: Harvard Univ. Press.
- Brickhouse, Thomas C.; Smith, Nicholas D.  
(2006): The socratic paradoxes. In: Benson, Hugh H. (2006) (Hg.): *A companion to Plato.* Malden: Blackwell, 263–277.
- Brink, David Owen  
(1989): *Moral realism and the foundations of ethics.* Cambridge: Cambridge Univ. Press.
- Broadie, Sarah  
(1991): *Ethics with Aristotle.* Oxford: Oxford Univ. Press.
- Bromwich, Danielle  
(2010): Clearing conceptual space for cognitivist motivational internalism. In: *Philosophical Studies* 148 (3), 343–367.
- Broome, John  
(1994): Discounting the future. In: *Philosophy & Public Affairs* 23 (2), 128–156.  
(1999): Normative requirements. In: *Ratio* 12 (4), 398–419.  
(2001): Are intentions reasons? And how should we cope with incommensurable values? In: Morris, Christopher; Ripstein, Arthur (2001) (Hg.): *Practical rationality and preference: essays for David Gauthier.* Cambridge Univ. Press, 98–120.
- Burnyeat, Myles Fredric  
(1980): Aristotle on learning to be good. In: Rorty 1980b, 69–92.
- Cane, Peter  
(2002): *Responsibility in law and morality.* Oxford: Hart.
- Chang, Ruth  
(1997) (Hg.): *Incommensurability, incomparability, and practical reason.* Cambridge: Harvard Univ. Press.  
(2001): Value pluralism. In: Smelser, Neil J.; Baltes, Paul B. (2001) (Hg.): *International encyclopedia of the social & behavioral sciences.* Amsterdam: Elsevier, 16139–16145.  
(2004): Can desires provide reasons for action? In: R. Jay Wallace, Philip Pettit, Samuel Scheffler und Michael Smith (2004) (Hg.): *Reason and value. Themes from the moral philosophy of Joseph Raz.* Oxford: Oxford Univ. Press, 56–90.
- Charles, David  
(1982-83): Rationality and irrationality. In: *Proceedings of the Aristotelian Society* 83, 191–212.  
(1984): *Aristotle's philosophy of action.* Ithaca: Cornell Univ. Press.  
(2009): Nicomachean ethics VII. 3: varieties of akrasia. In: Natali 2009a, 41–71.
- Chisholm, Robert  
(1966): Freedom and action. In: Lehrer, Keith (1966) (Hg.): *Freedom and determinism.* New York: Random House.
- Cholbi, Michael  
(2009): Moore's paradox and moral motivation. In: *Ethical Theory and Moral Practice* 12 (5), 495–510.
- Churchland, Paul M.  
(1970): The logical character of action-explanations. In: *The Philosophical Review* 79 (2), 214–236.
- Cooper, John M.  
(2009): Nicomachean ethics VII. 1–2: introduction, method, puzzles. In: Natali 2009a, 9–39.
- Corcilius, Klaus  
(2008a): *Streben und Bewegen. Aristoteles' Theorie der animalischen Ortsbewegung.* Berlin: de Gruyter (Quellen und Studien zur Philosophie, 79).  
(2008b): Akrasie bei Aristoteles: Die erste Aporie. In: Corcilius & Rapp 2008, 143–172.  
(2008c): Aristoteles' praktische Syllogismen in der zweiten Hälfte des 20. Jahrhunderts. In: Meixner, Uwe; Newen, Albert (2008) (Hg.): *Logical Analysis and History of Philosophy. Philosophiegeschichte und logische Analyse. Focus: the practical syllogism. Schwerpunkt: der praktische Syllogismus.* Guest Editors / Gastherausgeber Rapp, Christof; Brüllmann, Philipp. Paderborn: Mentis, 101–132.  
(2008d): Praktische Syllogismen bei Aristoteles. In: *Archiv für Geschichte der Philosophie* 90 (3), 247–297.

- Corcilius, Klaus; Rapp, Christof  
 (2008) (Hg.): Beiträge zur Aristotelischen Handlungstheorie. Stuttgart: Franz Steiner Verlag (Philosophie der Antike, 24).
- Cordner, Christopher  
 (1985): Jackson on weakness of will. In: *Mind* 94 (374), 273–280.
- Dahl, Norman O.  
 (1984): *Practical reason, Aristotle, and weakness of the will*. Minneapolis: Univ. of Minnesota Press.
- Dancy, Jonathan  
 (2000): *Practical reality*. Oxford: Oxford Univ. Press.
- Danto, Arthur C.  
 (1965): Basic actions. In: *American Philosophical Quarterly* 2 (2), 141–148.
- Darwall, Stephen L.  
 (1983): *Impartial reason*. Ithaca: Cornell Univ. Press.  
 (1992): Internalism and agency. In: *Philosophical Perspectives* 6, 155–174.
- Davidson, Donald  
 (1963): Actions, reasons, and causes. In: *Davidson 2001*, 3–19.  
 (1967): The logical form of action sentences. In: *Davidson 2001*, 105–122.  
 (1969a): How is weakness of the will possible. In: *Davidson 2001*, 21–42.  
 (1969b): The individuation of events. In: *Davidson 2001*, 163–180.  
 (1970): Mental events. In: *Davidson 2001*, 207–227.  
 (1971): Agency. In: *Davidson 2001*, 43–61.  
 (1973): Freedom to act. In: *Davidson 2001*, 63–81.  
 (1974): Belief and the basis of meaning. In: Davidson, Donald (2001): *Inquiries into truth and interpretation*. 2. ed. Oxford: Oxford University Press, 141–154.  
 (1976): Hempel on explaining action. In: *Davidson 2001*, 261–275.  
 (1978): Intending. In: *Davidson 2001*, 83–102.  
 (1982): Paradoxes of irrationality. In: *Davidson 2004*, 169–187.  
 (1985a): Replies to Essays I–IX. In: Vermazen & Hintikka 1985, 195–229.  
 (1985b): Reply to Quine on events. In: *Davidson 2001*, 305–311.  
 (1987): Problems in the explanation of action. In: *Davidson 2004*, 101–116.  
 (2001): *Essays on actions and events*. 2. ed. Oxford: Clarendon Press.  
 (2004): *Problems of rationality*. Oxford: Clarendon Press
- Davis, Lawrence H.  
 (1979): *Theory of action*. Englewood Cliffs: Prentice-Hall.
- Davis, Wayne A.  
 (2010): The causal theory of action. In: O'Connor & Sandis 2010, 32–39.
- Dodd, Dylan  
 (2009): Weakness of will as intention-violation. In: *European Journal of Philosophy* 17 (1), 45–59.
- Döring, Sabine A.  
 (2009): Kann Willensschwäche rational sein? In: Fürst, Martina; Gombocz, Wolfgang; Hiebaum, Christian (2009) (Hg.): *Gehirne und Personen. Beiträge zum 8. Internationalen Kongress der Österreichischen Gesellschaft für Philosophie in Graz. Band 1*. Frankfurt: Ontos-Verlag, 55–69.
- Dretske, Fred I.  
 (1999): Machines, plants and animals: the origins of agency. In: *Erkenntnis* 51 (1), 19–31.
- Dunn, Robert  
 (1987): *The possibility of weakness of will*. Indianapolis: Hackett.
- Elster, Jon  
 (1987): Introduction. In: Elster, Jon (1987) (Hg.): *The multiple self*. Cambridge: Cambridge Univ. Press, 1–34.  
 (2000): *Ulysses unbound. Studies in rationality, precommitment, and constraints*. Cambridge: Cambridge Univ. Press.
- Enç, Berent  
 (2004): Causal theories of intentional behavior and wayward causal chains. In: *Behavior and Philosophy* 32 (1), 149–166.
- Ferrero, Luca  
 (2005): The will: Interpersonal bargaining versus intrapersonal prediction. In: *Behavioral and Brain Sciences* 28 (5), 654–655.
- Fischer, John Martin  
 (1999): Recent work on moral responsibility. In: *Ethics* 110 (1), 93–139.
- Frankena, William  
 (1988): Hare on moral weakness and the definition of morality. In: *Ethics* 98 (4), 779–792.
- Frankfurt, Harry G.  
 (1969): Alternate possibilities and moral responsibility. In: *Frankfurt 2007*, 1–10.  
 (1971): Freedom of the will and the concept of a person. In: *Frankfurt 2007*, 11–25.

- (1978): The problem of action. In: Frankfurt 2007, 69–79.
- (2007): The importance of what we care about. Philosophical essays. 14. ed. Cambridge: Cambridge Univ. Press.
- Frederick, Shane; Loewenstein, George; O'Donoghue, Ted
- (2002): Time discounting and time preference: a critical review. In: *Journal of Economic Literature* 40 (2), 351–401.
- Frege, Gottlob
- (1993): *Begriffsschrift und andere Aufsätze*. Mit E. Husserls und H. Scholz' Anmerkungen hrsg. v. Ignacio Angelelli. 4. Nachdruck der 2. Auflage 1964. Hildesheim: Olms.
- Gardner, Sebastian
- (1993): *Irrationality and the philosophy of psychoanalysis*. Cambridge: Cambridge Univ. Press.
- Gardner, Thomas
- (2002): Socrates and Plato on the possibility of akrasia. In: *The Southern Journal of Philosophy* 40 (2), 191–210.
- Garrard, Eve; McNaughton, David
- (1998): Mapping moral motivation. In: *Ethical Theory and Moral Practice* 1 (1), 45–59.
- Geach, Peter T.
- (1965): Assertion. In: *The Philosophical Review* 74 (4), 449–465.
- Gerrig, Richard J.; Zimbardo, Philip G.
- (2008): *Psychologie*. 18. Aufl. München: Pearson Higher Education.
- Gibbard, Allan
- (1990): *Wise choices, apt feelings. A theory of normative judgement*. Oxford: Clarendon Press.
- Goldman, Alvin I.
- (1970): *A theory of human action*. Englewood Cliffs: Prentice-Hall.
- Gosling, Justin
- (1990): *Weakness of the will*. London: Routledge.
- Green, Leonard; Myerson, Joel
- (1993): Alternative frameworks for the analysis of self control. In: *Behavior and Philosophy* 21 (2), 37–47.
- Griffin, Richard; Dennett, Daniel
- (2005): Comparing apples to oranges: who does the framing? In: *Behavioral and Brain Sciences* 28 (05).
- Hardie, William Francis Ross
- (1968): *Aristotle's ethical theory*. 2. ed. Oxford: Clarendon Press 1980.
- Hare, Richard M.
- (1952): *The language of morals*. London, New York: Oxford Univ. Press.
- (1963): *Freedom and reason*. Oxford: Oxford Univ. Press.
- (1970): Meaning and speech acts. In: *The Philosophical Review* 79 (1), 3–24.
- (1981): *Moral thinking. Its levels, method, and point*. Oxford: Oxford Univ. Press.
- (1993): Objective prescriptions. In: Hare 1999, 1–18.
- (1998): 'Prescriptivism' In: *Routledge encyclopedia of philosophy*. London, New York: Routledge.
- (1999) (Hg.): *Objective prescriptions, and other essays*. Oxford, New York: Clarendon Press.
- Harrison, Glenn W.; Lau, Morten Igel
- (2005): Is the evidence for hyperbolic discounting in humans just an experimental artefact? In: *Behavioral and Brain Sciences* 28 (05).
- Hart, Herbert L. A.
- (2008) (Hg.): *Punishment and responsibility. Essays in the philosophy of law*. 2. ed. Oxford: Oxford Univ. Press.
- Heath, Joseph
- (1997): Foundationalism and practical reason. In: *Mind* 106 (423), 451–473.
- (2003): Practical irrationality and the structure of decision theory. In: Stroud & Tappolet 2003, 251–273.
- Heil, John
- (2004): *Philosophy of mind. A contemporary introduction*. 2. ed. New York: Routledge (Routledge contemporary introductions to philosophy).
- Hempel, Carl G.
- (1961–62): Rational action. In: *Proceedings and Addresses of the American Philosophical Association* 35, 5–23.
- Hempel, Carl G.; Oppenheim, Paul
- (1948): Studies in the logic of explanation. In: *Philosophy of Science* 15 (2), 135–175.
- Hill, Thomas E.
- (2005): Willensschwäche und Charakter. In: Spitzley 2005, 168–190. (= Hill, Thomas E. (1986): *Weakness of will and character*. In: *Philosophical Topics* 14 (2), 93–115).
- Hinchman, Edward S.
- (2009): Receptivity and the will. In: *Nous* 43 (3), 395–427.
- Höffe, Otfried
- (2010a): Ethik als praktische Philosophie – Methodische Überlegungen (I1, 1094a22–1095a13). In: Höffe 2010b, 13–38.
- (2010b) (Hg.): *Aristoteles: Nikomachische Ethik*. 3. Aufl. München: Oldenbourg Akademieverlag.

- Holton, Richard  
 (1999): Intention and weakness of will. In: *Journal of Philosophy* 96 (5), 241–262.  
 (2009): *Willing, wanting, waiting*. Oxford: Clarendon Press.
- Hooker, Brad; Streumer, Bart  
 (2004): Procedural and substantive practical rationality. In: Mele, Alfred R.; Rawling, Piers (2004) (Hg.): *The Oxford handbook of rationality*. Oxford: Oxford Univ. Press (Oxford Handbooks in Philosophy), 57–74.
- Horn, Christoph; Löhner, Guido  
 (2010a): Einleitung: Die Wiederentdeckung teleologischer Handlungserklärungen. In: Horn & Löhner 2010b, 7–45.  
 (2010b) (Hg.): *Gründe und Zwecke. Texte zur aktuellen Handlungstheorie*. Berlin: Suhrkamp.
- Hornsby, Jennifer  
 (1980): *Actions*. London: Routledge & Kegan Paul.
- Hughes, Paul M.  
 (2002): The logic of temptation. In: *Philosophia* 29, 89–110.
- Humberstone, Lloyd  
 (1992): Direction of fit. In: *Mind* 101 (401), 59–83.
- Hume, David  
 (1739): *A treatise of human nature*. Repr. with corr. Norton, David Fate; Norton, Mary J. (Hg.). Oxford: Oxford Univ. Press 2008.
- Hurka, Thomas  
 (1982): The speech act fallacy fallacy. In: *Canadian Journal of Philosophy* 1 (3), 509–526.
- Hurley, Susan L.  
 (1989): *Natural reasons. Personality and polity*. New York: Oxford Univ. Press 1992.
- ICD-10-GM  
 (2013): *Internationale statistische Klassifikation der Krankheiten und verwandter Gesundheitsprobleme. 10. Revision. German Modification*. Deutsches Institut für Medizinische Dokumentation und Information.
- Jackson, Frank  
 (1984): Weakness of will. In: *Mind* 93 (369), 1–18.
- Jones, Karen  
 (2003): Emotion, weakness of will, and the normative conception of agency. In: Hatzimoysis, Anthony (2003) (Hg.): *Philosophy and the emotions*. Cambridge: Cambridge Univ. Press, 181–200.
- Joyce, Richard  
 (2002): Expressivism and motivation internalism. In: *Analysis* 62 (4), 336–344.  
 (2009): Expressivism, motivation internalism, and Hume. In: Pigden, Charles R. (2009) (Hg.): *Hume on motivation and virtue*. Basingstoke: Palgrave Macmillan (Philosophers in depth), 30–56.
- Kalis, Annemarie; Mojzisch, Andreas; Schweizer, T. Sophie; Kaiser, Stefan  
 (2008): Weakness of will, akrasia, and the neuropsychiatry of decision making: An interdisciplinary perspective. In: *Cognitive, Affective, & Behavioral Neuroscience* 8 (4), 402–417.
- Kant, Immanuel  
 (1785): *Grundlegung zur Metaphysik der Sitten*. Kommentar v. Christoph Horn, Corinna Mieth und Nico Scarano. Frankfurt am Main: Suhrkamp 2007 (Suhrkamp Studienbibliothek, 2).
- Kauppinen, Antti  
 (2010): Moral judgment and volitional incapacity. In: Campbell, Joseph Keim; O'Rourke, Michael; Silverstein, Harry (2010) (Hg.): *Action, ethics, and responsibility*. Cambridge: MIT Press (Topics in contemporary philosophy), 235–258.
- Keil, Geert  
 (2000): *Handeln und Verursachen*. Frankfurt am Main: Klostermann (Philosophische Abhandlungen, 79).  
 (2007): *Willensfreiheit*. Berlin: de Gruyter (Grundthemen Philosophie).
- Kennett, Jeanette; Smith, Michael  
 (1994): Philosophy and commonsense: the case of weakness of will. In: Michael, Michaelis; O'Leary-Hawthorne, John (1994) (Hg.): *Philosophy in mind. The place of philosophy in the study of mind*. Dordrecht: Kluwer Academic Publ., 141–157.
- Kenny, Anthony John Patrick  
 (1975): *Will, freedom and power*. Oxford: Blackwell.  
 (1978): *The Aristotelian ethics*. Oxford: Clarendon Press.
- Korsgaard, Christine M.  
 (1986): Skepticism about practical reason. In: *The Journal of Philosophy* 83 (1), 5–25.
- Lear, Jonathan  
 (2009): Ruhelosigkeit, Phantasie und der Begriff des Geistes. In: *Deutsche Zeitschrift für Philosophie* 57 (1), 49–71.
- Lind, Robert C.  
 (2010) (Hg.): *Discounting for time and risk in energy policy*. Washington: Johns Hopkins Univ. Press.

- Liske, Michael-Thomas  
 (2008): Unter welchen Bedingungen sind wir für unsere Handlungen verantwortlich? In: Corcilius & Rapp 2008, 83–103.
- Little, Margaret Olivia  
 (1997): Virtue as knowledge: objections from the philosophy of mind. In: *Nous* 31 (1), 59–79.
- Loewenstein, George; Prelec, Drazen  
 (1992): Anomalies in intertemporal choice: evidence and an interpretation. In: *The Quarterly Journal of Economics* 107 (2), 573–597.
- Löhner, Guido  
 (2003): *Praktisches Wissen. Grundlagen einer konstruktiven Theorie menschlichen Handelns*. Paderborn: Mentis.  
 (2005): On ends and means. Constructive type theory as a guide for modeling in the theory of mind and action. In: *Mathematics and Social Sciences* 43 (171), 5–23.
- Mantel, Susanne  
 (2009): Motivierende Gründe und praktische Überlegung. In: *Zeitschrift für philosophische Forschung* 63 (2), 260–284.
- Martin-Löf, Per  
 (1984): *Intuitionistic type theory*. Naples: Bibliopolis.
- McIntyre, Alison  
 (1990): Is akratic action always irrational? In: Flanagan, Owen; Rorty, Amélie Oksenberg (1990) (Hg.): *Identity, character, and morality. Essays in moral psychology*. Cambridge: MIT Press, 379–400.
- Mele, Alfred R.  
 (1987a): *Irrationality. An essay on akrasia, self-deception, and self-control*. Oxford: Oxford Univ. Press.  
 (1987b): Intentional action and wayward causal chains: The problem of tertiary waywardness. In: *Philosophical Studies* 51 (1), 55–60.  
 (1987c): Recent work on self-deception. In: *American Philosophical Quarterly* 24 (1), 1–17.  
 (1992): Recent work on intentional action. In: *American Philosophical Quarterly* 29 (3), 199–217.  
 (1996): Internalist moral cognitivism and listlessness. In: *Ethics* 106 (4), 727–753.  
 (2010a): Weakness of will and akrasia. In: *Philosophical Studies* 150 (3), 391–404.  
 (2010b): Intention. In: O'Connor & Sandis 2010, 108–113.
- Mele, Alfred R.; Moser, Paul K.  
 (1994): Intentional action. In: *Nous* 28 (1), 33–68.
- Mill, John Stuart  
 (1974): *A system of logic. Ratiocinative and inductive. Being a connected view of the principles of evidence and the methods of scientific investigation. Books I–III*. Ed. by J. M. Robson. Toronto: Univ. of Toronto Press, Routledge & Kegan Paul.
- Millar, Alan  
 (2004): *Understanding people. Normativity and rationalizing explanation*. Oxford: Clarendon Press.
- Mitchell, Dorothy  
 (1982): Deviant causal chains. In: *American Philosophical Quarterly* 19 (4), 351–353.
- Morris, Michael  
 (2006): Akrasia in the Protagoras and the Republic. In: *Phronesis* 51 (3), 195–229.
- Moya, Carlos J.  
 (1990): *The philosophy of action. An introduction*. Cambridge: Polity Press.
- Müller, Jörn  
 (2003): Ergon und eudaimonia. Plädoyer für eine unifizierende Interpretation der ergon-Argumente in den aristotelischen Ethiken. In: *Zeitschrift für philosophische Forschung* 57 (4), 513–542.  
 (2009): *Willensschwäche in Antike und Mittelalter. Eine Problemgeschichte von Sokrates bis Johannes Duns Scotus*. Leuven: Leuven Univ. Press.
- Nagel, Thomas  
 (1970): *The possibility of altruism*. Princeton: Princeton Univ. Press.
- Natali, Carlo  
 (2009a) (Hg.): *Aristotle: Nicomachean ethics, Book VII. Symposium Aristotelicum*. Oxford: Oxford Univ. Press.  
 (2009b): *Nicomachean ethics VII. 5–6: beastliness, irascibility, akrasia*. In: Natali 2009a, 103–129.
- Nussbaum, Martha C.  
 (1984): Plato on commensurability and desire. In: *Proceedings of the Aristotelian Society, Supplement* 58, 55–80.  
 (1986): *The fragility of goodness. Luck and ethics in Greek tragedy and philosophy*. Rev. ed., reprint. Cambridge: Cambridge Univ. Press 2009.  
 (2001): *The Protagoras: a science of practical reasoning*. In: Millgram, Elijah (2001) (Hg.): *Varieties of practical reasoning*. Cambridge: MIT Press, 153–201.
- O'Connor, Timothy; Sandis, Constantine  
 (2010) (Hg.): *A companion to the philosophy of action*. Malden: Wiley-Blackwell (Blackwell companions to philosophy, 46).

- Owens, David  
 (2002): Epistemic akrasia. In: *The Monist* 85 (3), 381–397.
- Owens, Joseph  
 (1985): The akratic's "ultimate premise" in Aristotle. In: Wiesner, Jürgen (1985) (Hg.): *Aristoteles. Werk und Wirkung. Erster Band: Aristoteles und seine Schule*. Berlin: de Gruyter, 376–392.
- Parfit, Derek  
 (1997): Reasons and motivation. In: *Proceedings of the Aristotelian Society, Supplement* 71, 99–130.
- Peacocke, Christopher  
 (1985): Intention and akrasia. In: Vermazen & Hintikka 1985, 51–73.
- Pears, David  
 (1975): The appropriate causation of intentional basic actions. In: *Revista Hispanoamericana de Filosofía* 7 (20), 39–72.  
 (1982): How easy is akrasia? In: *Philosophia* 11 (1–2), 33–50.  
 (1984): *Motivated irrationality*. Oxford: Clarendon Press.
- Peijnenburg, Jeanne  
 (2005): Shaping your past selves. In: *Behavioral and Brain Sciences* 28 (5).
- Pendlebury, Michael  
 (2002): "Ought" judgments and motivation. In: *American Philosophical Quarterly* 39 (2), 183–196.
- Penner, Terry  
 (1992): Socrates and the early dialogues. In: Kraut, Richard (1992) (Hg.): *The Cambridge companion to Plato*. Cambridge: Cambridge Univ. Press, 121–169.  
 (1997): Socrates on the strength of knowledge: Protagoras 351B-357E. In: *Archiv für Geschichte der Philosophie* 79 (2), 117–149.
- Platon  
 (2008): *Politeia*. In: *Sämtliche Werke. Band 2. 32. Aufl.* Hrsg. v. Ursula Wolf. Reinbek bei Hamburg: Rowohlt-Taschenbuch-Verlag.  
 (2009): *Protagoras*. In: *Sämtliche Werke. Band 1. 31. Aufl.* Hrsg. v. Ursula Wolf. Reinbek bei Hamburg: Rowohlt-Taschenbuch-Verlag.
- Price, Anthony W.  
 (2006): Akrasia and self-control. In: Kraut, Richard (2006) (Hg.): *The Blackwell guide to Aristotle's nicomachean ethics*. Malden: Blackwell, 234–254.
- Pugmire, David  
 (1982): Motivated irrationality. In: *Proceedings of the Aristotelian Society, Supplement* 56, 179–196.
- Rachlin, Howard  
 (2000): *The science of self-control*. Cambridge: Harvard Univ. Press.  
 (2005): Problems with internalization. In: *Behavioral and Brain Sciences* 28 (5).
- Radoilska, Lubomira  
 (2012): Akrasia and ordinary weakness of will. In: *Tópicos* 43, 25–50.
- Ramsey, Frank  
 (1926): Truth and probability. In: Ramsey, Frank (1978): *Foundations. Essays in philosophy, logic, mathematics and economics*. Hg. v. David Mellor. London: Routledge & Kegan Paul (International library of psychology, philosophy and scientific method), 58–100.
- Ranta, Aarne  
 (1994): *Type-theoretical grammar*. Oxford: Clarendon Press (Indices, 1).
- Rapaport, Elizabeth  
 (1973): Explaining moral weakness. In: *Philosophical Studies* 24 (3), 174–182.
- Rapp, Christof  
 (2010): Freiwilligkeit, Entscheidung und Verantwortlichkeit (III 1–7). In: Höffe 2010b, 109–133.
- Rawls, John  
 (1971): *A theory of justice*. Revised edition. Cambridge: Belknap Press of Harvard Univ. Press 2003.
- Raz, Joseph  
 (1997): Incommensurability and agency. In: Chang 1997, 110–128.  
 (1999): Agency, reason, and the good. In: Raz, Joseph (1999) (Hg.): *Engaging reason. On the theory of value and action*. Oxford: Oxford Univ. Press, 22–45.
- Rebentisch, Juliane  
 (2009): Der Demokrat und seine Schwächen. Eine Lektüre von Platons *Politeia*. In: *Deutsche Zeitschrift für Philosophie* 57 (1), 15–36.
- Rebentisch, Juliane; Setton, Dirk  
 (2009): Schwerpunkt: Willensschwäche – Epistemologie und Politik irrationalen Handelns. In: *Deutsche Zeitschrift für Philosophie* 57 (1), 13–14.
- Resnik, Michael D.  
 (1987): *Choices. An introduction to decision theory*. 5. ed. Minneapolis: Univ. of Minnesota Press, 2000.

- Ribeiro, Brian  
 (2011): Epistemic akrasia. In: *International Journal for the Study of Skepticism* 1 (1), 18–25.
- Richman, Robert J.  
 (1980): Acrasia and practical reasoning. In: *Pacific Philosophical Quarterly* 61, 245–257.
- Ricken, Friedo  
 (1976): *Der Lustbegriff in der Nikomachischen Ethik des Aristoteles*. Göttingen: Vandenhoeck & Ruprecht.
- Ridge, Michael  
 (1998): Humean intentions. In: *American Philosophical Quarterly* 35 (2), 157–178.  
 (2006): Sincerity and expressivism. In: *Philosophical Studies* 131 (2), 487–510.
- Roberts, Robert C.  
 (1984): Will power and the virtues. In: *The Philosophical Review* 93 (2), 227–247.
- Robinson, Richard  
 (1969): Aristotle on akrasia. In: Höffe 2010b, 187–206.
- Rorty, Amélie Oksenberg  
 (1980a): Where does the akratic break take place? In: Rorty, Amélie (1988) (Hg.): *Mind in action. Essays in the philosophy of mind*. Boston: Beacon Press, 229–245.  
 (1980b) (Hg.): *Essays on Aristotle's ethics*. Berkeley: Univ. of California Press.  
 (1997): The social and political sources of akrasia. In: *Ethics* 107, 644–657.
- Rosati, Connie S.  
 (2008): 'Moral Motivation'. In: *The Stanford Encyclopedia of Philosophy* (Fall 2008 Edition). Hrsg. von Edward N. Zalta. URL = <<http://plato.stanford.edu/entries/moral-motivation/>>, zuletzt geprüft am 11.09.2013.
- Ross, William David  
 (1923): *Aristotle*. 6. ed. New York, London: Routledge 1995.
- Rothermund, Klaus; Eder, Andreas  
 (2011): *Motivation und Emotion*. 1. Aufl. Wiesbaden: VS Verlag für Sozialwissenschaften.
- Saarinen, Risto  
 (1994): *Weakness of the will in medieval thought. From Augustine to Buridan*. Leiden: Brill.
- Samuelson, Paul A.  
 (1937): A note on measurement of utility. In: *The Review of Economic Studies* 4 (2), 155–161.
- Santas, Gerasimos  
 (1966): Plato's Protagoras and explanations of weakness. In: *The Philosophical Review* 75 (1), 3–33.
- Sartre, Jean-Paul  
 (1943): *Das Sein und das Nichts. Versuch einer phänomenologischen Ontologie*. 14. Aufl. Hrsg. v. Traugott König und Vincent von Wroblewsky. Reinbek bei Hamburg: Rowohlt-Taschenbuch-Verlag 2008.
- Scanlon, Thomas  
 (1998): *What we owe to each other*. Cambridge: Belknap Press of Harvard Univ. Press.
- Schälike, Julius  
 (2002): *Wünsche, Werte und Moral. Entwurf eines handlungstheoretischen und ethischen Internalismus*. Würzburg: Königshausen & Neumann.  
 (2004): Willensschwäche und Selbsttäuschung. Über die Rationalität des Irrationalen und das Verhältnis von Evaluation und Motivation. In: *Deutsche Zeitschrift für Philosophie* 52 (3), 361–379.  
 (2005): Irrationale mentale Kausalität und praktische Normen a priori. In: *Zeitschrift für philosophische Forschung* 59 (1), 22–48.  
 (2009): Selbstkontrolle. Synchrone contra diachrone Analyse von motivationalem Zwang und Willensschwäche. In: *Studia Philosophica* 68, 255–273.
- Schroeder, Mark  
 (2007): *Slaves of the passions*. Oxford: Oxford Univ. Press.  
 (2008): *Being for. Evaluating the semantic program of expressivism*. Oxford: Oxford Univ. Press.
- Schroeder, Timothy  
 (2004): *Three faces of desire*. Oxford: Oxford Univ. Press (Philosophy of mind series).  
 (2009): 'Desire'. In: *The Stanford Encyclopedia of Philosophy* (Winter 2009 Edition). Hrsg. v. Edward N. Zalta. URL = <<http://plato.stanford.edu/archives/win2009/entries/desire/>>, zuletzt geprüft am 11.09.2013.
- Schueler, George Frederick  
 (1983): Akrasia revisited. In: *Mind* 92 (368), 580–584.  
 (1991): Pro-attitudes and direction of fit. In: *Mind* 100 (2), 277–281.
- Searle, John R.  
 (1962): Meaning and speech acts. In: *The Philosophical Review* 71 (4), 423–432.  
 (1969): *Speech acts. An essay in the philosophy of language*. Cambridge: Cambridge Univ. Press.  
 (1983): *Intentionality. An essay in the philosophy of mind*. Cambridge: Cambridge Univ. Press.
- Seel, Martin  
 (2002): Ein Lob der Willensschwäche. In: Seel, Martin (2002): *Sich bestimmen lassen. Studien zur theoretischen und praktischen Philosophie*. Frankfurt am Main: Suhrkamp, 227–245.

- Segvic, Heda  
 (2006): No one errs willingly: The meaning of Socratic intellectualism. In: Ahbel-Rappe, Sara; Kamtekar, Rachana (2006) (Hg.): *A companion to Socrates*. Malden: Blackwell, 171–184.
- Shafer-Landau, Russ  
 (2003): *Moral realism. A defence*. Oxford: Clarendon Press.
- Silverstein, Harry S.  
 (1970): Prescriptivism and akrasia. In: *Philosophical Studies* 21 (6), 81–85.
- Smith, Michael  
 (1987–88): Reason and desire. In: *Proceedings of the Aristotelian Society* 88, 243–358.  
 (1994): *The moral problem*. Oxford: Blackwell.  
 (1996): The Argument for internalism: Reply to Miller. In: *Analysis* 56 (3), 175–184.
- Sobel, Jordan Howard  
 (2001): On Michael Smith's internalism. In: *Erkenntnis* 54 (3), 345–373.
- Sobel, David; Copp, David  
 (2001): Against direction of fit accounts of belief and desire. In: *Analysis* 61 (1), 44–53.
- Spitzley, Thomas  
 (1992): Handeln wider besseres Wissen. Eine Diskussion klassischer Positionen. Berlin: de Gruyter (Quellen und Studien zur Philosophie, 30).  
 (2005) (Hg.): *Willensschwäche*. Übersetzungen von Joachim Schulte. Paderborn: Mentis (Mentis-Anthologien-Philosophie).  
 (2009): Weak-willed animals? In: *Rationality, Markets and Morals* 0, 95–108.
- Stocker, Michael  
 (1979): Desiring the bad: an essay in moral psychology. In: *The Journal of Philosophy* 76, 738–753.  
 (1984): Some structures for akrasia. In: *History of Philosophy Quarterly* 1 (3), 267–280.  
 (1990): Plural and conflicting values. Reprint. Oxford: Clarendon Press (Clarendon paperbacks).
- Stout, Rowland  
 (1980): Oblique causation and reasons for action. In: *Synthese* 43 (3), 351–367.  
 (2010): Deviant causal chains. In: O'Connor & Sandis 2010, 159–165.
- Stoutland, Frederick  
 (1980): Oblique causation and reasons for action. In: *Synthese* 43 (3), 351–367.  
 (1985): Davidson on intentional behavior. In: LePore, Ernest; McLaughlin, Brian P. (1985) (Hg.): *Actions and events. Perspectives on the philosophy of Donald Davidson*. Oxford: Basil Blackwell, 29–43.
- Strawson, Peter F.  
 (1962): Freedom and resentment. In: Strawson, Peter F. (2008): *Freedom and resentment and other essays*. London: Routledge, 1–28.
- Stroud, Sarah  
 (2003): Weakness of will and practical judgement. In: Stroud & Tappolet 2003, 121–146.
- Stroud, Sarah; Tappolet, Christine  
 (2003) (Hg.): *Weakness of will and practical irrationality*. Reprint. Oxford: Clarendon Press 2011.
- Svavarsdottir, Sigrun  
 (1999): Moral cognitivism and motivation. In: *The Philosophical Review* 108 (2), 161–219.
- Taylor, Charles  
 (1980): Plato, Hare and Davidson on akrasia. In: *Mind* 89 (356), 499–518.  
 (1984): Reply to Schueler on akrasia. In: *Mind* 93 (372), 584–586.
- Tenenbaum, Sergio  
 (1999): The judgment of a weak will. In: *Philosophy and Phenomenological Research* 59 (4), 875–911.  
 (2003): *Accidie*, evaluation, and motivation. In: Stroud & Tappolet 2003, 147–171.  
 (2007): *Appearances of the good. An essay on the nature of practical reason*. Cambridge: Cambridge Univ. Press.
- Thalberg, Irving  
 (1971): Acting against one's better judgment. In: Mortimore, Geoffrey (1971) (Hg.): *Weakness of will*. London: Macmillan, 233–246.  
 (1984): Do our intentions cause our intentional actions? In: *American Philosophical Quarterly* 21 (3), 249–260.
- Tieleman, Teun  
 (2009): Nicomachean ethics VII. 9 (1151b23) – 10: (in)continence in context. In: Natali 2009a, 173–182.
- Timmermann, Jens  
 (2000): Impulsivität und Schwäche. Die Argumentation des Abschnitts Eth. Nic. 1146b 31–1147b 19 im Licht der beiden Formen des Phänomens "Akrasia". In: *Zeitschrift für philosophische Forschung* 54 (1), 47–66.
- Urmson, James Opie  
 (1980): Aristotle's doctrine of the mean. In: Rorty 1980b, 157–170.
- Velleman, James David  
 (1985): Practical reflection. In: *The Philosophical Review* 94 (1), 33–61.  
 (1992): What happens when someone acts? In: *Mind* 101 (403), 461–481.

- (1993): The story of rational action. In: Velleman, James (2000) (Hg.): The possibility of practical reason. Oxford: Clarendon Press, 144–169.
- Vermazen, Bruce; Hintikka, Merrill B.  
 (1985) (Hg.): Essays on Davidson. Actions and events. Oxford: Oxford Univ. Press
- Vlastos, Gregory  
 (1969): Socrates on acrasia. In: Phoenix 23 (1), 71–88.
- von Neumann, John; Morgenstern, Oskar  
 (1944): Theories of games and economic behavior. 6. ed. Princeton: Princeton Univ. Press 1955.
- Walker, Arthur F.  
 (1989): The problem of weakness of will. In: Nous 23 (5), 653–676.
- Wallace, R. Jay  
 (1990): How to argue about practical reason. In: Mind 99, 355–385.  
 (2001): Normativity, commitment, and instrumental reason. In: Philosopher's Imprint 1 (4), 1–26.
- Watson, Gary  
 (1977): Skepticism about weakness of will. In: The Philosophical Review 86 (3), 316–339.
- Wedgwood, Ralph  
 (2007): The nature of normativity. Oxford: Oxford Univ. Press.
- Werner, Micha H.  
 (2006): Verantwortung. In: Düwell, Marcus; Hübenthal, Christoph; Werner, Micha H. (2006) (Hg.): Handbuch Ethik. 2., aktualisierte und erw. Aufl. Stuttgart: Metzler, 541–548.
- Wiggins, David  
 (1978): Weakness of will, commensurability, and the objects of deliberation and desire. In: Proceedings of the Aristotelian Society 79, 251–277.  
 (1980): Weakness of will, commensurability, and the objects of deliberation and desire. In: Rorty 1980b, 241–265. (leicht überarbeitete Version von Wiggins 1978)  
 (1997): Incommensurability: four proposals. In: Chang 1997, 52–66.
- Williams, Bernard  
 (1970): Deciding to believe. In: Williams, Bernard (1999): Problems of the self. Philosophical papers 1956–1972. Reprint. Cambridge: Cambridge Univ. Press, 139–151.  
 (1979): Internal and external reasons. In: Bernard Williams (1981): Moral luck. Philosophical papers 1973–1980. Cambridge: Cambridge Univ. Press, 101–113.  
 (1989): Internal reasons and the obscurity of blame. In: Williams, Bernard (1995): Making sense of humanity and other philosophical papers 1982–1993. Cambridge: Cambridge Univ. Press, 35–45.
- Wolf, Ursula  
 (1985): Zum Problem der Willensschwäche. In: Zeitschrift für philosophische Forschung 39 (1), 21–33.  
 (2007): Aristoteles' "Nikomachische Ethik". 2., durchges. Aufl. Darmstadt: Wissenschaftliche Buchgesellschaft.  
 (2010): Über den Sinn der Aristotelischen Mesoteslehre (II). In: Höffe 2010b, 83–108.
- Zangwill, Nick  
 (2008): Besires and the motivation debate. In: Theoria 74 (1), 50–59.