

## KANTS KOMPATIBILISMUS

von Tobias Rosefeldt

### 1. Kants Lügnerparadoxie

In der folgenden Passage aus der *Kritik der reinen Vernunft* beschreibt Kant das Dilemma, in das wir dadurch geraten, dass wir menschliche Handlungen einerseits als Teil der natürlichen Welt, andererseits aber als frei und moralisch bewertbar ansehen:

„[...] man [nehme] eine willkürliche Handlung, z.E. eine boshafte Lüge, durch die ein Mensch eine gewisse Verwirrung in die Gesellschaft gebracht hat, und die man zuerst ihren Bewegursachen nach, woraus sie entstanden, untersucht und darauf beurtheilt, wo sie sammt ihren Folgen ihm zugerechnet werden könne. In der ersten Absicht geht man seinen empirischen Charakter bis zu den Quellen desselben durch, die man in der schlechten Erziehung, übler Gesellschaft, zum Theil auch in der Bösartigkeit eines für Beschämung unempfindlichen Naturells aufsucht, zum Theil auf den Leichtsinns und Unbesonnenheit schiebt; wobei man denn die veranlassenden Gelegenheitsursachen nicht aus der Acht läßt. [...] Ob man nun gleich die Handlung dadurch bestimmt zu sein glaubt: so tadelt man nichts destoweniger den Thäter und zwar nicht wegen seines unglücklichen Naturells, nicht wegen der auf ihn einfließenden Umstände, ja sogar nicht wegen seines vorhergeführten Lebenswandels; denn man setzt voraus, man könne es gänzlich bei Seite setzen, wie dieser beschaffen gewesen, und die verflossene Reihe von Bedingungen als ungeschehen, diese That aber als gänzlich unbedingt in Ansehung des vorigen Zustandes ansehen, als ob der Thäter damit eine Reihe von Folgen ganz von selbst anhebe. [...] die Handlung wird seinem intelligibelen Charakter beigemessen, er hat jetzt, in dem Augenblicke, da er lügt, gänzlich Schuld; mithin war die Vernunft unerachtet aller empirischen Bedingungen der That völlig frei, und ihrer Unterlassung ist diese gänzlich beizumessen.“ (KrV A 554 f./ B 582 f.)<sup>1</sup>

Kants Position zu menschlicher Freiheit und ihrem Platz in der natürlichen Welt zeichnet sich dadurch aus, dass er beiden in dieser Passage genannten Einstellungen zu menschlichen Handlungen Rechnung tragen will und dafür argumentiert, dass sie miteinander vereinbar sind. Genauer kann diese Position anhand von drei Annahmen charakterisiert werden: eine *Annahme über die Determiniertheit menschlicher Handlungen durch zeitlich frühere Ereignisse*, eine *Annahme über die Möglichkeit transzendentaler Freiheit* und eine *Annahme über die Vereinbar-*

---

<sup>1</sup> Kants Schriften werden hier zitiert nach: Immanuel Kant, *Gesammelte Schriften*, hg. von der Preußischen [später: Deutschen] Akademie der Wissenschaften, Berlin 1900 ff. (abgekürzt „AA“). Die Angabe von Stellen aus der *Kritik der reinen Vernunft* folgt dabei der Paginierung der Originalausgaben der ersten („A“) und zweiten Auflage („B“).

keit der ersten beiden Annahmen unter der Voraussetzung, dass man den transzendentalen Idealismus als Theorie über Raum und Zeit akzeptiert.

Eine Formulierung der ersten Annahme, die dem heutigen Verständnis von Determinismus recht nahe kommt, findet sich z.B. in Kants Zustimmung zu der Vermutung, dass menschliche „Handlungen als Erscheinungen durch und durch mit anderen Erscheinungen nach beständigen Naturgesetzen im Zusammenhange ständen und von ihnen als ihren Bedingungen abgeleitet werden könnten und also mit diesen in Verbindung Glieder einer einzigen Reihe der Naturordnung ausmachen“ (KrV A 539/B 567). Eine weitere Belegstelle ist seine Aussage, dass „jede Handlung, die in einem Zeitpunkte vorgeht, unter der Bedingung dessen, was in der vorhergehenden Zeit war, nothwendig sei“ (AA V 94). Der von Kant bevorzugte Ausdruck für diese Art von Determiniertheit von Handlungen durch zeitlich frühere Zustände lautet „Prädeterminismus“ (vgl. AA VI 49 f.). Ich werde Kants Annahme, dass die genannte Determiniertheit in der raum-zeitlichen Welt besteht, aber dem heutigen Sprachgebrauch folgend als *Determinismusthese* bezeichnen. Man kann die folgendermaßen formulieren:

#### *Determinismusthese*

Für jede in der Zeit stattfindende menschliche Handlung  $H$  gibt es einen zeitlich vor  $H$  liegenden Naturzustand  $Z$  und Naturgesetze  $N_1 \dots N_n$ , so dass gilt: Es ist unmöglich, dass  $Z$  existiert und  $N_1 \dots N_n$  gelten, aber  $H$  nicht stattfindet.

Obwohl die genannten Textstellen eine recht deutliche Sprache sprechen, ist die These, dass Kant einen Determinismus im Sinne diese Annahme vertreten hat, in der Literatur allerdings immer wieder bestritten worden.<sup>2</sup> In diesem Band zum Beispiel argumentiert Jochen Bojanowski dagegen, dass Kant Determinist im Sinne der Determinismusthese war.<sup>3</sup> Er stützt sich dabei auf Kants Bemerkungen in den Antinomien, dass es keine vollständige Reihe der hinreichenden Ursachen für ein bestimmtes Ereignis geben kann, man also auch nicht davon ausgehen kann, dass menschliche Handlungen durch den vorhergehenden Weltverlauf vollständig bestimmt sind. Ich halte dieses Argument für nicht überzeugend, weil der Determinismus gar nicht anzunehmen braucht, dass es eine *vollständige* Reihe von Ursachen für eine bestimmte Handlung geben muss, damit diese Handlung durch einen vorherigen Naturzustand vollständig kausal determiniert ist. In der obigen Formulierung der Determinismusannahme und den Textstellen, auf denen sie beruht, ist allein davon die Rede, dass es für jede Handlung einen zeitlich früheren Zustand gibt, der

<sup>2</sup> Vgl. etwa Michael Wolff: *Freiheit und Determinismus bei Kant* (erscheint in: B. Tuschling & W. Euler (Hg.), *Kants Metaphysik der Sitten. Philosophische und editorische Probleme*. Berlin).

<sup>3</sup> Vgl. neben seinem Beitrag zu diesem Band auch Jochen Bojanowski: *Kants Theorie der Freiheit. Rekonstruktion und Rehabilitierung*. Berlin 2006.

zusammen mit den Naturgesetzen die Handlung determiniert. Ob dieser Zustand selbst verursacht wird, vor allem aber, ob man beim Rückgang zu immer weiteren Ursachen jemals an ein Ende kommt oder nicht, ist dabei irrelevant.<sup>4</sup> Generell scheinen mir Zweifel an jeder Art von Interpretation angebracht, die darauf hinausläuft, dass Kants einen Determinismus im Rahmen seines Systems entweder gar nicht vertreten hat oder aber darunter etwas hinsichtlich der Vereinbarkeit mit Freiheit ‚Harmloseres‘ verstanden hat, als man das in der gegenwärtigen Debatte tut. Passagen wie die folgende machen meines Erachtens unmissverständlich klar, dass Kant der Meinung war, dass sich transzendente Freiheit auch mit dem radikalsten Determinismus Laplacescher Prägung verträgt:

„Man kann also einräumen, daß, wenn es für uns möglich wäre, in eines Menschen Denkungsart, so wie sie sich durch innere sowohl als äußere Handlungen zeigt, so tiefe Einsicht zu haben, daß jede, auch die mindeste Triebfeder dazu uns bekannt würde, imgleichen alle auf diese wirkende äußere Veranlassungen, man eines Menschen Verhalten auf die Zukunft mit Gewißheit, so wie eine Mond- oder Sonnenfinsterniß ausrechnen könnte und dennoch dabei behaupten, daß der Mensch frei sei.“ (AA V 99)<sup>5</sup>

Die zweite oben genannte Annahme, diejenige über menschliche Freiheit, findet sich z.B. in der folgenden Passage:

„[...] jede Handlung unangesehen des Zeitverhältnisses, darin sie mit anderen Erscheinungen steht, ist die unmittelbare Wirkung des intelligibelen Charakters der reinen Vernunft, welche mithin frei handelt, ohne in der Kette der Naturursachen durch äußere oder innere, aber der Zeit nach vorhergehende Gründe dynamisch bestimmt zu sein [...]“ (KrV A 553/B 581)

Wesen, die zu derjenigen Art von Handlung fähig sind, die wie die hier erwähnte Handlung der Vernunft nicht durch etwas anderes – z.B. einen zeitlich früheren Zustand – verursacht

---

<sup>4</sup> Vgl. zu diesem Kritikpunkt auch *Derk Peereboom: Kant on Transcendental Freedom*. – In: *Philosophy and Phenomenological Research* 73 (2006). 537-567, Fußnote 24.

<sup>5</sup> Jochen Bojanowski argumentiert in seinem Beitrag zu diesem Band dafür, dass Kant an dieser Stelle keinen Prädeterminismus, d.h. keine Annahme der Determiniertheit menschlicher Handlungen durch zeitlich frühere Ereignisse vertritt. Auch wenn dieses exegetische Manöver für die Sonnenfinsternispassage erfolgreich sein sollte, scheint sie mir als generelle Strategie nicht erfolgversprechend. Der Kontext der Stelle aus der Religionsschrift, in der Kant zwischen Determinismus und Prädeterminismus unterscheidet (AA VI 49 f.), macht deutlich, dass er selbst auch den Prädeterminismus für die raum-zeitliche Welt annehmen will und gerade diese Annahme als philosophische Herausforderung betrachtet. An dieser und verschiedenen anderen Stellen, zum Beispiel aus der *Kritik der praktischen Vernunft* (vgl. AA V 94 f. und AA V 97), wird klar, dass es gerade die Determiniertheit von Handlungen durch vergangene Ereignisse, die nicht mehr in der Gewalt des Handelnden liegen, ist, die Kants Grund für die Annahme darstellen, dass menschliche Freiheit nur dann zu retten ist, wenn Menschen nicht an sich in der Zeit existieren.

ist, bezeichnet Kant als „transzendental frei“ (vgl. z.B. A 446/ B 474). Seine Annahme über menschliche Freiheit kann man also folgendermaßen formulieren:

*Freiheitsthese*

Menschen sind transzendental frei, d.h. manche ihrer in der Zeit stattfindenden Handlungen sind die Wirkungen eines Handelns ihrer Vernunft, das selbst unverursacht ist.

Kant ist nicht der Meinung, dass wir theoretisches Wissen von der Wahrheit dieser These erlangen können. Da er die Wahrheit der Freiheitsthese für eine notwendige Voraussetzung dafür hält, dass Menschen zu moralisch richtigem Handeln verpflichtet sind und für moralisch falsches Handeln getadelt werden dürfen, misst er ihr aber den selben Grad und die selbe Art von Gewissheit zu wie der Überzeugung, dass diese beiden Sachverhalte bestehen (vgl. z.B. *Kritik der praktischen Vernunft*, AA V 3 ff.).

Kants Vereinbarkeitsthese schließlich lautet folgendermaßen:

*Vereinbarkeitsthese*

Die Determinismusthese und die Freiheitsthese können dann und nur dann beide wahr sein, wenn der transzendente Idealismus wahr ist, d.h. wenn raum-zeitliche Eigenschaften Gegenständen nur als Erscheinungen zukommen, diese Gegenstände an sich selbst aber nicht in Raum und Zeit existieren. Menschen können dann als Dinge an sich transzendental frei sein, obwohl sie und ihre Handlungen als Erscheinungen kausal durch frühere Weltzustände determiniert sind.

Der erste, „dann, wenn“-Teil der Vereinbarkeitsthese findet sich z.B. in folgendem Ausschnitt aus der *Kritik der praktischen Vernunft*:

„[Der] Widerstreit zwischen Naturnothwendigkeit und Freiheit [ist] kein wahrer Widerstreit [...], da ein und dasselbe handelnde Wesen als Erscheinung [...] eine Causalität in der Sinnenwelt hat, die jederzeit dem Naturmechanismus gemäß ist, in Ansehung derselben Begebenheit aber, so fern sich die handelnde Person zugleich als Noumenon betrachtet (als reine Intelligenz, in seinem nicht der Zeit nach bestimmbar Dasein), einen Bestimmungsgrund jener Causalität nach Naturgesetzen, der selbst von allem Naturgesetze frei ist, enthalten könne.“ (AA V 114)

Der zweite, „nur dann, wenn“-Teil der Vereinbarkeitsthese kommt zum Ausdruck, wenn Kant klarstellt, dass sich transzendente Freiheit und Determinismus nur dann vereinbaren lassen, wenn man zugesteht, dass Dinge an sich nicht in der Zeit existieren, und zwar des-

wegen, weil jede Handlung in der Zeit durch zeitlich vorhergehende Ereignisse determiniert ist, auf die der Handelnde zum Zeitpunkt der Handlung keinen Einfluss mehr hat. Als Beleg kann wieder ein Ausschnitt aus der *Kritik der praktischen Vernunft* dienen:

„Nimmt man nun die Bestimmungen der Existenz der Dinge in der Zeit für Bestimmungen der Dinge an sich selbst [...], so läßt sich die Nothwendigkeit im Causalverhältnisse mit der Freiheit auf keinerlei Weise vereinigen; sondern sie sind einander contradictorisch entgegengesetzt. Denn aus der ersteren folgt: daß eine jede Begebenheit, folglich auch jede Handlung, die in einem Zeitpunkte vorgeht, unter der Bedingung dessen, was in der vorhergehenden Zeit war, nothwendig sei. Da nun die vergangene Zeit nicht mehr in meiner Gewalt ist, so muß jede Handlung, die ich ausübe, durch bestimmende Gründe, die nicht in meiner Gewalt sind, nothwendig sein, d.i. ich bin in dem Zeitpunkte, darin ich handle, niemals frei. [...] in jedem Zeitpunkte stehe ich doch immer unter der Nothwendigkeit, durch das zum Handeln bestimmt zu sein, was nicht in meiner Gewalt ist, und [...] meine Causalität [wäre] also niemals Freiheit.“ (AA V 94 f.; vgl. auch AA V 97)

Auf Grund der Vereinbarkeitsthese sollte man Kants Position in der Freiheitsdebatte meines Erachtens als Kompatibilismus bezeichnen. Dies ist in letzter Zeit zwar bestritten worden und zwar mit dem Argument, dass Kant eine Freiheitsauffassung hat, die normalerweise von Inkompatibilisten vertreten wird und zeitgenössische Kompatibilisten Freiheit und Determinismus in der Regel deswegen für kompatibel halten, weil sie eine Analyse des Freiheitsbegriffes favorisieren, der Kant nicht zustimmen würde.<sup>6</sup> Aber „Kompatibilismus“ ist dem Wortsinn nach ein Ausdruck für die These, dass menschliche Freiheit und Determinismus vereinbar sind, und nicht ein Ausdruck für eine bestimmte Auffassung darüber, weshalb und unter Inkaufnahme welcher weiteren theoretischen Kosten sie vereinbar sind.<sup>7</sup> Die theoretischen Kosten von Kants Kompatibilismus sind zweifellos hoch. Determinismus lässt sich für Kant mit Freiheit – im für ihn einzig relevanten Sinne von transzendentaler Freiheit – nur dann vereinbaren, wenn man akzeptiert, dass Menschen nur als Erscheinungen, nicht aber an sich selbst, in der Zeit existieren und dass sie – oder, wie es in dem obigen Zitat heißt, „ihre Vernunft“ – als Dinge an sich außerhalb der Zeit handeln

<sup>6</sup> Vgl. z.B. Simon Shengjian Xie: *What Is Kant: A Compatibilist or an Incompatibilist? A New Interpretation of Kant's Solution to the Free Will Problem.* – In: *Kant-Studien* 100 (2009). 53–76. Zu Kants Kritik an dem Freiheitsbegriff, der den gewöhnlichen Versionen des Kompatibilismus zu Grunde liegt, vgl. z.B. *Kritik der praktischen Vernunft* AA V 95 ff.

<sup>7</sup> Allen Wood hat Kants Position als „compatibilism between compatibilism and incompatibilism“ bezeichnet, um den Unterschied zu gewöhnlichen Versionen des Kompatibilismus zu markieren (vgl. *Kant's Compatibilism.* – In: *Allen Wood (Ed.): Self and Nature in Kant's Philosophy.* Ithaca 1984. 73–101). Am treffendsten finde ich die ebenfalls leicht paradoxe Bezeichnung „libertarischer Kompatibilismus“ von Wolfgang Ertl, weil sie deutlich macht, dass Kant eine libertarische Freiheitskonzeption mit dem Naturdeterminismus vereinbaren will (vgl. *Wolfgang Ertl: 'Ludwig' Molina and Kant's Libertarian Compatibilism.* – Erscheint in: *Matthias Kaufmann et al. (Ed.): A Companion to Luis de Molina.* Köln, Leiden. New York 2012).

können. Zumindest behauptet Kant dies gemäß der Lesart, die ich als *Standardinterpretation von Kants Kompatibilismus* bezeichnen möchte.

Auch diese Standardinterpretation ist in der Literatur immer wieder bestritten worden, und man hat versucht, Kants Kompatibilismus so zu interpretieren, dass dessen theoretische Kosten weniger spektakulär ausfallen. Ein Beispiel für diese Strategie ist Hud Hudsons Versuch, Kants Position als frühe Version eines anomischen Monismus à la Davidson zu interpretieren.<sup>8</sup> Kants These wäre dann, dass Handlungen und die Entscheidungen, die zu ihnen führen, zwar unter der Beschreibung als Teile der physikalischen Welt, nicht aber unter der Beschreibung als mentale und rationale Zustände unter ein deterministisches Naturgesetz fallen. Auch diesen Versuch halte ich für exegetisch unhaltbar.<sup>9</sup> Erstens meint Kant, dass eine transzendental freie Handlung selbst unverursacht sein muss; eine Handlung, die unter einer Beschreibung unter ein Kausalgesetz fällt und unter einer anderen Beschreibung nicht, ist aber nicht unverursacht. Zweitens macht er unmissverständlich klar, dass es unter der Voraussetzung des Determinismus keine Freiheit geben könnte, wenn alle Handlungen in der Zeit stattfinden würden, weil in diesem Fall jede Handlung in der Zeit durch etwas determiniert ist, „was zur vergangenen Zeit gehört und nicht mehr in seiner [des Handelnden; T.R.] Gewalt“ ist.<sup>10</sup> Die Annahmen des anomischen Monismus implizieren aber sicher nicht, dass Ereignisse unter irgendeiner Beschreibung außerhalb der Zeit stattfinden.

Die wichtigste Motivation dafür, nach Alternativen zur Standardinterpretation von Kants Kompatibilismus zu suchen, besteht sicher darin, dass diese Interpretation Kant eine Konzeption unterstellt, die viele für so abwegig halten, dass sie ihm diese entweder als wohlwollende Interpreten oder als Philosophen, die an einer plausiblen Rekonstruktion der eigentlich wichtigen Einsichten Kants interessiert sind, nicht zuschreiben wollen. Diese Motivation ist durchaus nachvollziehbar. Kants Kompatibilismus impliziert gemäß der Standardinterpretation nicht nur die problematische Unterscheidung zwischen Erscheinungen und Dingen an sich sowie die Annahme, dass Dinge an sich selbst nicht in Raum und Zeit existieren; diese Interpretation beinhaltet darüber hinaus auch, dass in der Zeit stattfindende Handlungen in Kants Konzeption auf eine bestimmte Weise überdeterminiert sind: sie sind einerseits durch frühere Naturzustände und die Naturgesetze determiniert, andererseits hängen sie kausal von einem transzendental freien und zeitlosen Handeln der Vernunft ab. Das scheint aber zu implizieren, dass unser transzendental freies Handeln nur dann einen Einfluss auf unsere Handlungen in der Zeit haben kann, wenn es auch einen Einfluss auf

---

<sup>8</sup> *Hud Hudson: Kant's Compatibilism*. Ithaca, London 1994.

<sup>9</sup> Vgl. dazu auch die Kritik in *Wolfgang Ertl: Hud Hudson: Kant's Compatibilism*. – In: *Kant-Studien*. 90 (1999). 371–384 und in *Pereboom: Kant on transcendental freedom* (Anm. 4).

<sup>10</sup> Vgl. da obige Zitat aus der *Kritik der praktischen Vernunft* AA V 98 f., sowie AA V 101: „In der That: wären die Handlungen des Menschen, so wie sie zu seinen Bestimmungen in der Zeit gehören, nicht bloße Bestimmungen desselben als Erscheinung, sondern als Dinges an sich selbst, so würde die Freiheit nicht zu retten sein“.

die Geltung der Naturgesetze oder auf den bisherigen Weltverlauf hat, die zusammen über das Stattfinden der zeitlichen Handlung entscheiden. Diese Konsequenz scheint aber absurd zu sein.

Ich möchte in diesem Beitrag versuchen, die Standardinterpretation von Kants Kompatibilismus gegen den eben skizzierten Einwand zu verteidigen. Dazu werde ich zum einen untersuchen, welche Konsequenzen Kants Kompatibilismus gemäß der hier vertretenen Standardinterpretation für den Zusammenhang zwischen freiem zeitlosem Handeln, determiniertem zeitlichem Handeln und den dieses Handeln determinierenden Faktoren tatsächlich hat; zum anderen werde ich dafür argumentieren, dass keine dieser Konsequenzen sachlich so absurd ist, dass sie einen hinreichenden Grund dafür darstellen würde, diese Form des Kompatibilismus' nicht selbst zu vertreten oder sie Kant nicht zuzuschreiben (Abschnitte 3–5). Im nächsten Abschnitt werde ich zunächst allerdings noch etwas genauer auf einige Besonderheiten von Kants Konzeption freien Handelns und seiner Determinismusauffassung eingehen, die für meine Darstellung und Verteidigung von Kants Kompatibilismus relevant sind (Abschnitt 2). Nicht verteidigen werde ich Kant gegen den Vorwurf, dass sein Kompatibilismus nur dann funktioniert, wenn man die Unterscheidung zwischen in der Zeit existierenden Erscheinungen und außerhalb der Zeit existierenden Dingen an sich hinnimmt, und dass eben diese Unterscheidung schlicht nicht akzeptabel ist. Das ist natürlich ein triftiger Grund dafür, Kants Kompatibilismus abzulehnen. Aber mir reicht es, hier dafür zu argumentieren, dass es auch der *einzigste* triftige Grund ist und dass Kants Kompatibilismus also zumindest nicht abwegiger ist als Kants transzendentaler Idealismus im allgemeinen.<sup>11</sup>

## 2. Empirischer und intelligibler Charakter, empirische und intelligible Tat

Kant formuliert die Determinismusthese, die Freiheitsthese und die Vereinbarkeitsthese häufig mit Rekurs auf den sog. „empirischen“ und den sog. „intelligiblen Charakter“ eines Menschen, so zum Beispiel in der folgenden Passage:

„Es muß aber eine jede wirkende Ursache einen *Charakter* haben, d.i. ein Gesetz ihrer Causalität, ohne welches sie gar nicht Ursache sein würde. Und da würden wir an einem Subjecte der Sinnenwelt erstlich einen *empirischen Charakter* haben, wodurch seine Handlungen als Erscheinungen durch und durch mit anderen Erscheinungen nach beständigen Naturgesetzen im Zusammenhange ständen und von ihnen als ihren Bedingungen abgeleitet

---

<sup>11</sup> Dass Kants transzendentaler Idealismus im allgemeinen nicht so abwegig ist, wie es *prima facie* den Anschein hat, habe ich zu zeigen versucht in: *Dinge an sich und sekundäre Qualitäten*. – In: Jürgen Stolzenberg (Hg.): *Kant in der Gegenwart*. Berlin, New York 2007. 167–209.

werden könnten und also mit diesen in Verbindung Glieder einer einzigen Reihe der Naturordnung ausmachen. Zweitens würde man ihm noch einen *intelligibelen Charakter* einräumen müssen, dadurch es zwar die Ursache jener Handlungen als Erscheinungen ist, der aber selbst unter keinen Bedingungen der Sinnlichkeit steht und selbst nicht Erscheinung ist. Man könnte auch den ersteren den Charakter eines solchen Dinges in der Erscheinung, den zweiten den Charakter des Dinges an sich selbst nennen. [...] Nach seinem empirischen Charakter würde [...] dieses Subject als Erscheinung allen Gesetzen der Bestimmung nach der Causalverbindung unterworfen sein; und es wäre so fern nichts, als ein Theil der Sinnenwelt, dessen Wirkungen, so wie jede andere Erscheinung aus der Natur unausbleiblich abfließen. [...] Nach dem intelligibelen Charakter desselben aber [...] würde dasselbe Subject dennoch von allem Einflusse der Sinnlichkeit und Bestimmung durch Erscheinungen freigesprochen werden müssen; und da in ihm, so fern es *Noumenon* ist, nichts *geschieht*, keine Veränderung, welche dynamische Zeitbestimmung erheischt, mithin keine Verknüpfung mit Erscheinungen als Ursachen angetroffen wird, so würde dieses thätige Wesen so fern in seinen Handlungen von aller Naturnothwendigkeit, als die lediglich in der Sinnenwelt angetroffen wird, unabhängig und frei sein. [...] So würde denn Freiheit und Natur, jedes in seiner vollständigen Bedeutung, bei eben denselben Handlungen, nachdem man sie mit ihrer intelligibelen oder sensibelen Ursache vergleicht, zugleich und ohne allen Widerstreit angetroffen werden.“ (KrV A 539–41/B 567–9)

Auch an anderen Stellen nennt Kant bei der Formulierung seiner Determinismusannahme den empirischen Charakter des Menschen und die äußeren Umstände der jeweiligen Handlung als diejenigen Faktoren, die menschliche Handlungen determinieren. So schreibt er, dass „alle Handlungen des Menschen in der Erscheinung aus seinem empirischen Charakter und den mitwirkenden anderen Ursachen nach der Ordnung der Natur bestimmt“ sind (KrV A550/B578) und dass „jede [willkürliche Handlung] im empirischen Charakter des Menschen vorher bestimmt [ist], ehe noch als sie geschieht“ (KrV A 553/B 581).<sup>12</sup> Unter dem empirischen Charakter einer kausal wirksamen Substanz versteht Kant „ein Gesetz ihrer Causalität“, also etwas, das festlegt, auf welche Weise die Substanz unter gegebenen Umständen handelt, d.h. Zustände in sich selbst oder anderen Substanzen bewirkt.<sup>13</sup> Kants Determinismusannahme kann man also auch folgendermaßen formulieren:

---

<sup>12</sup> Im Anfangszitat dieses Beitrags wird das, was ich hier „äußere Umstände der Handlung“ nenne, nicht wie in diesem Zitat „mitwirkende andere Ursachen“, sondern „veranlassenden Gelegenheitsursachen“ genannt (vgl. KrV A 554/B 582). Dafür, dass mit diesen Formulierungen dasselbe gemeint ist, vgl. die *Metaphysikvorlesung L<sub>2</sub> nach Pölitx* (AA XXVIII 572).

<sup>13</sup> Der hier virulente Handlungsbegriff ist nicht der heute übliche, sondern derjenige der metaphysischen Tradition zu Kants Zeit. Handeln tun beliebige Substanzen diesem Begriff zufolge genau dann, wenn sie vermöge einer eigenen Kraft einen Zustand in sich selbst oder in einer anderen Substanz hervorrufen (vgl. z.B. *Metaphysikvorlesung L<sub>2</sub> nach Pölitx*, AA XXVIII 564).

*Determinismusthese\**

Für jede in der Zeit stattfindende menschliche Handlung  $H$  gibt es einen empirischen Charakter  $C_E$  des Handelnden und zeitlich vor  $H$  liegende Handlungsumstände  $U$ , so dass gilt: Es ist unmöglich, dass der Handelnde  $C_E$  hat und die Handlungsumstände  $U$  vorliegen, aber  $H$  nicht stattfindet.<sup>14</sup>

Unter dem intelligiblen Charakter eines Menschen ist in Analogie zum empirischen Charakter etwas zu verstehen, das festlegt, auf welche Weise der Mensch unter gegebenen Umständen als Ding an sich wirksam ist. Da der Mensch als Ding an sich nicht in der Zeit existiert, lässt es sein intelligibler Charakter zu, dass er handelt bzw. etwas verursacht, ohne dass es für diese Handlung selbst eine Ursache geben muss – dass er also im Sinne der Freiheitsthese transzendental frei ist. Zu den Dingen, die der Mensch auf Grund seines intelligiblen Charakters bewirkt, gehören auch seine in der Zeit stattfindenden empirischen Handlungen, die allerdings auch durch seinen empirischen Charakter und die Handlungsumstände festgelegt sind. Laut Kant ist diese Doppeldetermination deswegen möglich, weil es ein Abhängigkeitsverhältnis zwischen intelligiblem und empirischem Charakter gibt: Kant nennt den intelligiblen Charakter die „transscendentale Ursache“ des empirischen Charakters (A 546/B 574) und sagt, dass der empirische Charakter „im intelligiblen Charakter [...] bestimmt“ sei (A 551/B 579) und als „sinnliches Zeichen desselben“ angesehen werden könne (A 546/B 574).

Mit der Abhängigkeit des empirischen Charakters vom intelligiblen will Kant die moralische Zurechenbarkeit empirischer Handlungen erklären. Zwar ist jede empirische Handlung durch den empirischen Charakter und Handlungsumstände determiniert, aber weil „ein anderer intelligibler Charakter [...] einen andern empirischen gegeben haben“ würde (A 556/B 584), hätte eine moralisch schlechte Handlung wie eine Lüge auch unterlassen werden können, dann nämlich, wenn der intelligente Charakter des Lügenden ein anderer gewesen wäre. Diese Erklärung ist allerdings nur dann plausibel, wenn man die moralische Zurechenbarkeit des intelligiblen Charakters selbst für unproblematisch hält. Doch dies ist keineswegs selbstverständlich: Nur weil der intelligente Charakter einer ist, der dem Menschen als Ding an sich zukommt, ist er nicht unbedingt etwas, das der Mensch zu verantworten hat und das ihn schuldig macht. Schließlich scheint es viele anderen Dinge zu geben, die keine vernünftig handelnden Wesen sind, und denen man als Dingen an sich einen intelligiblen Charakter zuschreiben kann, *ohne* sie deswegen für irgendetwas, das aus diesem intelligiblen Charakter folgt, verantwortlich zu machen.

Kant hat dieses Problem spätestens seit der *Kritik der praktischen Vernunft* gesehen und darauf mit der Annahme reagiert, dass ein Mensch deswegen für seinen intelligiblen Charakter

---

<sup>14</sup> Ich werde weiter unten etwas dazu sagen, wie diese Version der Determinismusthese mit der ursprünglichen zusammenhängt.

verantwortlich sind, weil er selbst die Ursache dafür ist, dass er diesen Charakter hat. Er schreibt dort, dass wir einen Menschen deswegen für eine gesetzeswidrige Handlung verantwortlich machen, weil sie „zu einem einzigen Phänomen seines Charakters [gehört], *den er sich selbst verschafft*, und nach welchem er sich als einer von aller Sinnlichkeit unabhängigen Ursache die Causalität jener Erscheinungen selbst zurechnet“ (AA V 98; Hervorhebung von mir, T.R.). Mit dem Motiv der selbständigen Aneignung eines intelligiblen Charakters nimmt Kant dabei die antike Konzeption der Lebenswahl auf und kann sich zudem auf unsere alltäglichen moralischen Überzeugungen berufen, denen zufolge der Verweis darauf, dass sich eine bestimmte moralisch schlechte Handlung durch den schlechten Charakter des Handelnden erklären lässt, nicht als Entschuldigung zählt.

Zu voller Entfaltung bringt Kant die Konzeption einer selbständigen Aneignung eines intelligiblen Charakters erst in seiner Schrift *Die Religion innerhalb der Grenzen der bloßen Vernunft*, die eine umfassende Auseinandersetzung mit der Frage enthält, wie es böse Handlungen geben kann, d.h. Handlungen, die moralisch falsch, aber dennoch frei und zurechenbar sind. Das Problem, mit dem sich Kant konfrontiert sieht, besteht darin, dass solche Handlungen einerseits nicht daraus resultieren können, dass der Mensch vernunftgesteuert handelt, seine Vernunft aber selbst „boshaft“ ist, d.h. ihm unmoralische Anweisungen gibt, andererseits aber auch nicht dadurch erklärt werden können, dass der Mensch beim Handeln vollständig von seinen sinnlichen Neigungen bestimmt ist, denn in diesem Fall wäre seine Handlung nicht mehr frei.<sup>15</sup> Kants Lösung dieses Dilemmas besteht in der Annahme, dass der Mensch sowohl das moralische Gesetz als auch das Prinzip der Selbstliebe „als Triebfedern in seine Maxime“ aufnimmt, er aber „die Triebfeder der Selbstliebe und ihre Neigungen zur Bedingung der Befolgung des moralischen Gesetzes macht“ (AA VI 36), d.h. es sich zur obersten Maxime macht, dem Sittengesetz nur dann zu folgen, wenn dies nicht zu sehr seiner Selbstliebe widerspricht.<sup>16</sup> An dieser Stelle nimmt Kant nun seine Kon-

---

<sup>15</sup> Vgl. AA VI 35: „Um also einen Grund des Moralisch-Bösen im Menschen anzugeben, enthält die Sinnlichkeit zu wenig; denn sie macht den Menschen, indem sie die Triebfedern, die aus der Freiheit entspringen können, wegnimmt, zu einem bloß Thierischen; eine vom moralischen Gesetze aber freisprechende, gleichsam boshafte Vernunft (ein schlechthin böser Wille) enthält dagegen zu viel, weil dadurch der Widerstreit gegen das Gesetz selbst zur Triebfeder (denn ohne alle Triebfeder kann die Willkür nicht bestimmt werden) erhoben und so das Subject zu einem teuflischen Wesen gemacht werden würde. – Keines von beiden aber ist auf den Menschen anwendbar.“

<sup>16</sup> Alternativ könnte man Kant so verstehen, dass wir es uns zur obersten Maxime gemacht haben, dem Sittengesetz nur dann zu folgen, wenn deren Gebot sich *mit dem der Selbstliebe deckt*. Ich halte die im Haupttext gewählte Formulierung allerdings für der Sache nach plausibler. Sie passt zudem sehr gut zu dem Zusammenhang, den Kant herstellt zwischen der falschen Unterordnung der Triebfedern in der obersten Maxime und der These, dass „ein jeder Mensch [...] seinen Preis [hat], für den er sich weggiebt“ (AA VI 38). Dieser Formulierung zufolge besteht der menschliche Hang zum Bösen nicht darin, dass wir nur dann das moralische Gebotene tun, wenn wir es ohnehin auch aus Neigung getan hätten, sondern darin, dass wir dem Sittengesetz nicht bedingungslos folgen, d.h. es für jeden von uns einen Grad der Verletzung unserer Eigeninteressen gibt, bei dem wir nicht mehr dem moralischen Gebot folgen würden. Das ließe zu, dass Menschen zwar alle gleichermaßen böse sind, weil es überhaupt einen Punkt gibt, an dem der Konflikt zwischen Pflicht und Neigung zu Gunsten der Neigung ausgeht, aber sich dennoch darin unterscheiden, wo bei ihnen dieser Punkt liegt.

zeption eines sich selbst „verschafften“ intelligiblen Charakters des Menschen wieder auf: Der intelligible Charakter des Menschen – d.h. das „Gesetz seiner Kausalität“ – zeichnet sich nämlich gerade durch diese Rangordnung seiner Triebfedern in der obersten Handlungsmaxime aus (vgl. AA VI 37, 47 und 21 f.). Kant argumentiert wieder dafür, dass die genannte Rangordnung nur dann die Möglichkeit zurechenbarer böser Handlungen verständlich machen kann, wenn sie selbst als das Resultat einer freien Handlung verstanden wird, weil „nichts sittlich- (d.i. zurechnungsfähig-) böse [ist], als was unsere eigene That ist“ (AA VI 31).

Was Kant nun zum ersten Mal in aller Deutlichkeit betont ist, dass seine Erklärung zuschreibbarer böser Handlungen nur dann funktioniert, wenn er zwei verschiedene Arten von freien Handlungen, d.h. (in Kants Terminologie) von „Taten“ oder „facta“ unterscheidet.<sup>17</sup> Er schreibt:

„Es kann aber der Ausdruck von einer Tat überhaupt sowohl von demjenigen Gebrauch der Freiheit gelten, wodurch die oberste Maxime (dem Gesetz gemäß oder zuwider) in die Willkür aufgenommen, als auch von demjenigen, da die Handlung selbst (ihrer Materie nach, d.i. die Objekte der Willkür betreffend) jener Maxime gemäß ausgeübt werden. [...] Jene ist intelligible Tat, bloß durch Vernunft ohne alle Zeitbestimmungen erkennbar; diese sensibel, empirisch, in der Zeit gegeben (factum phaenomenon). Die erste heißt nun vornehmlich in Vergleichung mit der zweiten ein bloßer Hang und angeboren, [...] weil wir davon, warum in uns das Böse gerade die oberste Maxime verderbt habe, obgleich dieses unsere eigene That ist, eben so wenig weiter eine Ursache angeben können, als von einer Grundeigenschaft, die zu unserer Natur gehört.“ (AA VI 31 f.)

Es ist auffällig, dass Kant die oberste Handlungsmaxime an dieser Stelle durch zwei Merkmale charakterisiert, die in einer gewissen Spannung zueinander stehen: Einerseits soll sie so etwas wie eine grundlegende Eigenschaft des menschlichen Willens sein, die man als „angeboren“ bezeichnen könnte und mit einer „Grundeigenschaft unserer Natur“ vergleichen kann. Andererseits soll sie das Resultat einer freien Handlung der Aneignung sein, weil nur so die empirischen Handlungen, die dieser Maxime gemäß ausgeübt werden, zurechenbar sind. Diese Spannung kommt auch an anderen Stellen der Religionsschrift zum Ausdruck, so zum Beispiel an der folgenden:

„Die eine oder die andere [d.h. böse oder gute; T.R.] Gesinnung als angeborne Beschaffenheit von Natur haben, bedeutet hier auch nicht, daß sie von dem Menschen, der sie hegt,

---

Der eine lügt erst dann, wenn er ansonsten gefoltert würde, der andere, weil er sich dadurch einen Dokortitel erschwindeln kann.

<sup>17</sup> Zum Verständnis von „Tat“ und „factum“ als „freie Handlung“ vgl. z.B. AA VI 223 und *Metaphysikvorlesung L<sub>2</sub> nach Pölitz*, AA XXVIII 565.

gar nicht erworben, d.i. er nicht Urheber sei; sondern daß sie nur nicht in der Zeit erworben sei (daß er eines oder das andere von Jugend auf sei immerdar). Die Gesinnung [...] muß auch durch freie Willkür angenommen worden sein, denn sonst könnte sie nicht zugerechnet werden. Von dieser Annehmung kann nun nicht wieder der subjective Grund oder die Ursache erkannt werden (obwohl darnach zu fragen unvermeidlich ist: weil sonst wiederum eine Maxime angeführt werden müßte, in welche diese Gesinnung aufgenommen worden, die eben so wiederum ihren Grund haben muß). Weil wir also diese Gesinnung, oder vielmehr ihren obersten Grund nicht von irgend einem ersten Zeit-Actus der Willkür ableiten können, so nennen wir sie eine Beschaffenheit der Willkür, die ihr (ob sie gleich in der That in der Freiheit gegründet ist) von Natur zukommt.“ (AA VI 25)

Es ist klar, aus welchen theoretischen Motiven Kant den intelligiblen Charakter bzw. die Gesinnung eines Menschen mit der darin festgeschriebenen verkehrten Rangordnung seiner Triebfedern als das Resultat einer zeitlosen und willentlichen intelligiblen Tat verstehen will. Es ist in der Literatur aber immer wieder moniert worden, dass der Begriff einer solchen intelligiblen Tat letztlich unverstündlich bleibt.<sup>18</sup> Erstens scheint angesichts dessen, was Kant über diese Tat sagt, unklar, wie sie aus „freier Willkür“ geschehen, d.h. eine willentliche Handlung sein kann. Zweitens ist fraglich, ob es der Begriff des Handelns nicht ohnehin ausschließt, dass es Handlungen gibt, die nicht zu einer bestimmten Zeit stattfinden oder eine bestimmte Dauer haben.

Auf das erste Problem kann ich im Rahmen dieses Beitrags nicht eingehen und möchte nur darauf hinweisen, dass Kant selbst unter der Willentlichkeit der intelligiblen Tat etwas anderes verstanden haben muss, als dasjenige, was gemäß seiner Handlungstheorie einer gewöhnlichen Handlung zukommt, wenn sie willentlich geschieht.<sup>19</sup> Wie Kant selbst schreibt, kann die intelligible Tat, mit der man sich auf die oberste Maxime festlegt, nicht das Resultat einer praktischen Überlegung oder bewussten Entscheidung sein, weil dies selbst bereits das Vorhandensein einer obersten Handlungsmaxime voraussetzen würde. Diese Feststellung beantwortet allerdings natürlich nicht die Frage, wie eine positive Charakterisierung der Willentlichkeit der intelligiblen Tat aussehen sollte.<sup>20</sup>

Das zweite Problem jedoch, das im Zuge der Auseinandersetzung mit der Kompatibilismusproblematik das drängendere ist, lässt sich meines Erachtens ausräumen, und zwar

---

<sup>18</sup> Für eine Darstellung der Probleme, die der Begriff der intelligiblen Tat mit sich bringt, vgl. z.B. *Marcus Willaschek: Praktische Vernunft. Handlungstheorie und Moralbegründung bei Kant.* Stuttgart, Weimar 1992. 149–167.

<sup>19</sup> Vgl. zum Begriff ‚gewöhnlicher‘ Handlungen aus Willkür z.B. *Die Metaphysik der Sitten*, AA VI 213.

<sup>20</sup> Die systematisch aussichtsreichste Antwort auf diese Frage scheint mir zu sein, dass die Willentlichkeit der intelligiblen Tat einer Bestimmung der obersten Maxime nichts weiter ist als ein Teilaspekt der Willentlichkeit der empirischen Handlungen, die aus der Befolgung dieser Maxime resultieren. Um diese Annahme zu rechtfertigen, müsste man dafür argumentieren, dass empirische Handlungen nur dann als solche gelten können, die man wirklich ausführen will, wenn man auch die Handlungsmaximen haben will, aus deren Befolgung die Handlungen resultieren.

dann, wenn man genauer darauf achtet, welcher Handlungsbegriff Kants Aussagen zu Grunde liegt. Wie bereits erläutert<sup>21</sup> deckt sich Kants Handlungsbegriff nicht mit demjenigen, was man heute „Handlung“ nennen würde, sondern entstammt der vorkantischen metaphysischen Tradition, in der Handeln („actio“) einfach das Gegenteil von Leiden („passio“) ist. Dieser Handlungsbegriff lässt es zu, beliebigen – also zum Beispiel auch unbelebten – Substanzen das Vermögen zu handeln zuzusprechen; sie handeln genau dann, wenn sie vermöge ihrer eigenen Kraft Ursache eines Zustands ihrer selbst oder einer anderen Substanz sind. Dass Kant nicht mehr als dies unter einer Handlung versteht, lässt sich anhand seiner Metaphysikvorlesungen belegen (vgl. z.B. *Metaphysikvorlesung L<sub>2</sub> nach Pölitz*, AA XXVIII 564), er weist darauf aber zum Beispiel auch in der *Kritik der reinen Vernunft* hin, wenn er zu Beginn der transzendentalen Deduktion schreibt, dass jemand, der wissen will, wie sich die sogenannten Prädikabilien – zu denen der Begriff der Handlung gehört – aus den von ihm explizit behandelten Kategorien ergeben, am besten „die Ontologischen Lehrbücher zur Hand nimmt“ (A 82/B 108). In dem für Kant maßgeblichen dieser Lehrbücher – d.h. in Baumgartens *Metaphysica* – wird der Begriff der Handlung in § 210 folgendermaßen definiert:

„Handeln (Agieren, Tätigsein) ist eine Veränderung des Zustands, und ganz allgemein die Verwirklichung eines Akzidens in der Substanz durch deren eigene Kraft; Leiden ist eine Veränderung des Zustands, und ganz allgemein die Verwirklichung eines Akzidens in der Substanz durch fremde Kraft.“<sup>22</sup>

Interessant an dieser Definition ist nun, dass sie Handeln im allgemeinsten Sinne nicht an eine Zustandsveränderung koppelt, sondern als „Verwirklichung eines Akzidens in der Substanz durch deren eigene Kraft“ beschreibt. Dadurch ist der Handlungsbegriff kompatibel damit, dass etwas unter ihn fällt, das nicht in der Zeit stattfindet, denn anders als der Begriff der Zustandsveränderungen ist der Begriff der Verwirklichung („actio“) eines Akzidens in einer Substanz kein wesentlich zeitlicher, zumindest dann nicht, wenn man mit Kant annimmt, dass Dinge überhaupt Eigenschaften haben können, ohne sie zu einem bestimmten Zeitpunkt zu haben. Die intelligible Tat als zeitlose Handlung (in Kants Sinne) zu beschreiben, ist also beileibe kein begrifflicher Widerspruch.

Dass eine Substanz handelt, wenn diese Verwirklichung eine „aus eigener Kraft“ ist, kann man in Anschluss an § 197 der *Metaphysica* verstehen, in dem Baumgarten eine Kraft als *Grund* dafür definiert, dass einer Substanz bestimmte Akzidenzien inhärieren. Eine Substanz handelt also genau dann, wenn sie selbst die Ursache dafür ist, dass sie oder eine an-

<sup>21</sup> Vgl. oben Anm. 13.

<sup>22</sup> Zitiert nach *Alexander Gottlieb Baumgarten: Metaphysica*. Übersetzt, eingeleitet und herausgegeben von Günter Gawlick und Lothar Kreimendahl. Stuttgart 2011. 137.

dere Substanz sich in einem bestimmten Zustand befinden. Und ihr Handeln ist transzendental frei, wenn es dafür, dass sie eine solche Ursache ist, keine weitere Ursache in einer anderen Substanz gibt. Dass der Mensch sich seinen intelligiblen Charakter und die damit verbundene oberste Handlungsmaxime in einer intelligiblen freien Handlung selbst verschafft, bedeutet also nicht mehr, als dass die Tatsache, dass er diesen Charakter hat, ontologisch fundamental und in nichts als dem jeweiligen Menschen selbst gegründet ist, d.h. nicht dadurch weiter erklärt – und entschuldigt – werden kann, dass ein anderes Ding auf bestimmte Weise beschaffen ist.<sup>23</sup>

Fassen wir die wichtigsten Punkte von Kants Erklärung der Freiheit und Zuschreibbarkeit böser Handlungen wie der im Anfangszitat dieses Beitrags erwähnten boshaften Lüge noch einmal zusammen: Wir müssen eine transzendental freie und zeitlose intelligible Tat annehmen, die darin besteht, dass sich der Mensch als Ding an sich einen bestimmten intelligiblen Charakter aneignet, d.h. es sich zur obersten Maxime macht, den Forderungen des moralischen Gesetzes nur dann zu folgen, wenn diese nicht zu sehr dem Prinzip seiner Selbstliebe widersprechen. Diese Aneignung der Charakters durch den Menschen darf dabei nicht als zeitliches Abfolgeverhältnis, sondern soll als eines der ontologischen Fundierung verstanden werden. Neben der zeitlosen freien Handlung gibt es auch in der Zeit stattfindende freie Handlungen (z.B. Lügen). Diese sind deswegen frei und moralisch zurechenbar, weil sie nach der genannten obersten Maxime ausgeführt werden und weil sie nur deswegen stattfinden, weil der Mensch denjenigen intelligiblen Charakter hat, den er sich „verschafft“ hat. Die moralische Zurechenbarkeit einer bösen Handlung ist deswegen damit vereinbar, dass diese Handlung durch den empirischen Charakter des Menschen und die äußeren Umstände seiner Handlung determiniert ist, weil der empirische Charakter

---

<sup>23</sup> Es ist natürlich eine alles andere als leicht zu beantwortende Frage, was man unter ontologischer Fundierung, die nicht in einem Kausalverhältnis zeitlich aufeinander folgender Ereignisse besteht, verstehen soll und was es genau heißen soll, dass eine bestimmte Beschaffenheit einer Substanz in dieser Substanz allein gegründet ist. Vor dem Hintergrund der Renaissance von Konzeptionen nicht ereignis-kausaler ontologischer Fundierung in der gegenwärtigen Debatte über „grounding“ kann man Kants Konzeption allerdings nicht mehr einfach deswegen als absurd abtun, weil er überhaupt eine solche Art von Fundierung annimmt. Was die Frage nach den historischen Quellen des Zusammenhangs zwischen transzendentaler Freiheit und ontologischer Fundiertheit des Charakters einer Substanz in dieser Substanz selbst betrifft, hat Wolfgang Ertl auf die so erstaunliche wie erhellende Parallele zwischen Kants Konzeption und derjenigen des scholastischen Philosophen Luis de Molina hingewiesen (vgl. *Wolfgang Ertl: 'Ludwig' Molina and Kant's Libertarian Compatibilism.* – Erscheint in: *Matthias Kaufmann et al. (Ed.): A Companion to Luis de Molina.* Köln, Leiden. New York 2012). In Molinas Freiheitskonzeption werden durch die Freiheit eines Menschen Wahrheiten darüber konstituiert, wie dieser Mensch in gegebenen Situationen handeln würde, und zwar Wahrheiten, die einerseits kontingent sind, aber andererseits trotzdem nicht vom Willen Gottes abhängen, sondern von diesem bei der Schöpfung als gegeben hingenommen werden müssen. Gott als Primärursache der Welt kann für Molina deswegen nur im Zusammenwirken mit den Menschen als Sekundärursachen bestimmen, was in der Welt geschieht. Es ist nicht schwer, in Kants These, dass der Charakter eines Menschen auf seiner intelligiblen Tat beruht und in nichts anderem als dem Menschen selbst gegründet ist, ein Echo dieser Molinistischen Freiheitskonzeption zu hören. Ertl zeigt ferner, dass es keineswegs ausgeschlossen ist, dass Kant tatsächlich von Molinas Lehre beeinflusst wurde.

eines Menschen seinerseits von seinem intelligiblen Charakter abhängt. Alles in allem kann der Lügner also deswegen für seine boshafte Lüge moralisch verantwortlich gemacht werden, weil das folgende komplexe Abhängigkeitsverhältnis besteht: Hätte die intelligente Tat der falschen Triebfederunterordnung nicht stattgefunden, hätte der Mensch einen anderen intelligiblen Charakter gehabt (nämlich einen, der in der Überordnung des Moralprinzips über das der Selbstliebe besteht). Hätte er diesen anderen intelligiblen Charakter gehabt, hätte er auch einen anderen empirischen Charakter gehabt (nämlich einen, der in der Situation, in der es zu der Lüge kam, dazu geführt hätte, dass der Mensch diese unterlässt). Und hätte der Mensch einen solchen anderen empirischen Charakter gehabt, dann hätte er nicht gelogen.

Ich werde in Abschnitt 5 auf die Frage zurückkommen, ob der Zusammenhang zwischen intelligibler Handlung, intelligentem Charakter, empirischem Charakter und empirischer Handlung durch die eben erwähnte kontrafaktische Abhängigkeit bereits hinreichend beschrieben ist. Zuvor möchte ich mich allerdings wie angekündigt mit dem Einwand auseinandersetzen, dass schon die genannten kontrafaktischen Abhängigkeiten Kant auf die *prima facie* absurde These verpflichten, dass wir durch unsere intelligente Handlung nicht nur Einfluss auf unsere Handlungen in der Zeit nehmen, sondern darüber hinaus auch entweder auf die Geltung der Naturgesetze oder auf den Weltverlauf, der zu den zeitlichen Handlungen geführt hat.

### 3. *Are we free to break the laws?*

Ich werde den genannten Einwand im folgenden anhand des Beispiels der im Eingangszitat erwähnten „boshafte Lüge“ darstellen. Im folgenden soll „L“ für diese in der Zeit stattfindende Handlung des Lügens stehen und „H“ für diejenige transzendental freie Handlung der Vernunft, mit der sich der Lügner seinen intelligiblen Charakter verschafft hat, auf Grund deren wir ihn für seine Lüge verantwortlich machen. Ferner soll „ $N_1 \dots N_n$ “ für die in der Erscheinungswelt geltenden Naturgesetze stehen und „Z“ für einen Naturzustand vor L, der zusammen mit  $N_1 \dots N_n$  L determiniert. Wir können das folgende gültige Argument formulieren:

- (1) Es ist unmöglich, dass Z existiert und  $N_1 \dots N_n$  gelten, aber L nicht stattfindet. (*Determinismusthese*)
- (2) Hätte H nicht stattgefunden, dann hätte auch L nicht stattgefunden. (*Annahme über die kontrafaktische Abhängigkeit von L und H*)
- (K<sub>0</sub>) Also: Hätte H nicht stattgefunden, dann hätte entweder Z nicht existiert oder aber  $N_1 \dots N_n$  nicht gegolten.

An dieser Stelle kann man auf mindestens drei Weisen fortfahren: Wer der Meinung ist, dass die Vergangenheit nicht von unseren freien Handlungen abhängen sollte, aber mit Kant an (1) und (2) festhalten will, wird die Konsequenz (K<sub>1</sub>) ziehen:

(K<sub>1</sub>) Hätte *H* nicht stattgefunden, dann hätten *N*<sub>1</sub>...*N*<sub>*n*</sub> nicht gegolten.

Nennen wir jemanden, der (K<sub>1</sub>) für eine akzeptable Konsequenz hält, einen „*altered laws*“-Kompatibilisten“.

Wer meint, dass die Geltung der Naturgesetze nicht von unseren freien Handlungen abhängen, aber mit Kant an (1) und (2) festhalten will, wird die Konsequenz (K<sub>2</sub>) ziehen:

(K<sub>2</sub>) Hätte *H* nicht stattgefunden, dann hätte *Z* nicht existiert.

Wer diese Konsequenz für akzeptabel hält, soll im folgenden „*altered past*-Kompatibilist“ heißen. Wer schließlich meint, dass weder die Vergangenheit noch die Geltung der Naturgesetze von unseren freien Handlungen abhängen, wird das Argument für eine Widerlegung von Kants Kompatibilismus halten.

Verteidiger von Kants Kompatibilismus haben sowohl für den ‚altered laws‘-Kompatibilismus als auch für den ‚altered past‘-Kompatibilismus argumentiert. Ich werde in diesem Abschnitt den ‚altered laws‘-Kompatibilismus und im nächsten Abschnitt den ‚altered past‘-Kompatibilismus diskutieren und dafür argumentieren, dass beide Positionen sehr viel weniger abwegig sind, als sie auf den ersten Blick erscheinen mögen.

Der ‚altered laws‘-Kompatibilismus behauptet, dass die in unserer Welt geltenden Naturgesetze von unseren transzendental freien Handlungen abhängen.<sup>24</sup> Eric Watkins hat dafür argumentiert, dass diese Behauptung im Rahmen von Kants Kausalitätsauffassung sehr viel weniger abwegig ist, als man es auf den ersten Blick vermutet. Um seine Verteidigung des ‚altered laws‘-Kompatibilismus zu verstehen, muss man sich vor Augen führen, dass man Prämisse (1) des oben genannten Arguments unter Rückgriff auf die in Abschnitt 2 erläuterte alternative Version der Determinismussthese auch anders formulieren könnte. Das obige Argument für die Konklusion (K<sub>0</sub>) kann dann folgendermaßen reformuliert werden („*U*“ stehe an dieser Stelle für die äußeren Umstände der Lüge, „*C<sub>E</sub>*“ für den tatsächlichen empirischen Charakter des Lügenden):

---

<sup>24</sup> Der ‚altered laws‘-Kompatibilismus wird unter anderem vertreten in *Wolfgang Ertl: Schöpfung und Freiheit. Ein kosmologischer Schlüssel zu Kants Kompatibilismus*. – In: Norbert Fischer (Hg.): *Kants Metaphysik und Religionsphilosophie*. Hamburg 2004. 43–76, in *Erik Watkins: Kant and the Metaphysics of Causality*. Cambridge 2005, und in *Pereboom: Kant on transcendental freedom* (Anm. 4).

- (1\*) Es ist unmöglich, dass die Handlungsumstände  $U$  bestehen und der Handelnde  $C_E$  hat, aber  $L$  nicht stattfindet. (*Determinismusthese\**)
- (2) Hätte  $H$  nicht stattgefunden, dann hätte auch  $L$  nicht stattgefunden. (*Annahme über die kontrafaktische Abhängigkeit von  $L$  und  $H$* )
- ( $K_0^*$ ) Also: Hätte  $H$  nicht stattgefunden, dann hätten entweder die Handlungsumstände  $U$  nicht bestanden oder der Handelnde hätte nicht  $C_E$  gehabt.

Das Argument, dass es unplausibel ist,  $H$  einen Einfluss auf die Handlungsumstände von  $L$  zuzusprechen, spricht dann für die folgende Variante der Konsequenz ( $K_1$ ):

( $K_1^*$ ) Hätte  $H$  nicht stattgefunden, dann hätte der Handelnde nicht  $C_E$  gehabt.

Wie hängen die ursprünglichen Versionen der Determinismusthese, des Arguments für ( $K_0^*$ ) und der Konsequenz ( $K_1$ ) mit ihren \*-Varianten zusammen? Wenn man Eric Watkins' Interpretation von Kants Theorie der Kausalität folgt, gibt es einen sehr engen Zusammenhang zwischen ihnen.<sup>25</sup> Watkins argumentiert dafür, dass in Kants Kausalitätsauffassung die Relata der Kausalitätsrelation nicht Ereignisse sind, sondern als Wirkungen Ereignisse und als Ursachen Substanzen fungieren. Substanzen sind insofern Ursachen, als sie in bestimmten Umständen eine bestimmte „Kausalität“ zeigen, wobei Kant den bereits erwähnten empirischen Charakter einer Substanz als das „Gesetz ihrer Kausalität“ versteht, d.h. als etwas, das festlegt, welche Wirkungen die Substanz in bestimmten Umständen hervorbringt.<sup>26</sup> Die speziellen Naturgesetze in einer Welt ergeben sich dann aus der Summe der empirischen Charaktere der die Welt bildenden Substanzen. Insofern gilt die Determinismusthese in ihrer ersten Variante, weil sie in ihrer (\*)-Variante gilt. Und aus der Wahrheit von ( $K_1^*$ ) folgt die von ( $K_1$ ), denn wenn der eigene empirische Charakter ein anderer gewesen wäre, wären auch die Naturgesetze, die sich ja aus der Gesamtheit der empirischen Charaktere ergeben, andere gewesen.

Watkins argumentiert selbst dafür, dass seine Interpretation von Kants Kausalitätskonzeption eine plausible Lesart von Kants Kompatibilismus möglich macht. Die These, dass die Naturgesetze andere gewesen wären, wenn unsere transzendental freie Handlung eine andere gewesen wäre, verliert ihren Schrecken, wenn man bedenkt, dass dies deswegen der Fall gewesen wäre, weil die Naturgesetze in einer Welt allgemein über den empirischen Charakteren der Substanzen in dieser Welt supervenieren, und die empirischen Charaktere von Substanzen in deren intelligiblen Charakteren gegründet sind.

<sup>25</sup> Zum folgenden vgl. *Watkins: Metaphysics of Causality* (Anm. 24), Kap. 4 und 5.

<sup>26</sup> Vgl. die oben zitierte Passage aus KrV A 539–41/B 567–9; Watkins spricht in seiner Rekonstruktion eher von Naturen als von Charakteren, meint mit beiden Ausdrücken aber dasselbe.

Watkins' Variante des ‚altered laws‘-Kompatibilismus ist allerdings mit einem schwerwiegenden Problem konfrontiert. Das Problem ergibt sich daraus, dass Kant nicht nur annimmt, dass zeitliche Handlungen wie Lügen kausal durch frühere Naturzustände und den empirischen Charakter des Handelnden determiniert sind, sondern dass auch der empirische Charakter selbst durch frühere Ereignisse in der Welt festgelegt wird. In dem Anfangszitat hatte es geheißen, dass man dann, wenn man eine Lüge „ihren Bewegursachen nach, woraus sie entstanden, untersucht“, den „empirischen Charakter [des Lügners; T.R.] bis zu den Quellen desselben durch[geht], die man in der schlechten Erziehung, übler Gesellschaft, zum Theil auch in der Bösartigkeit eines für Beschämung unempfindlichen Naturells aufsucht, zum Theil auf den Leichtsinns und Unbesonnenheit schiebt“ (KrV A 554/ B 582). Das bedeutet: Bei der kausalen Erklärung einer Handlung wie einer Lüge behandeln wir den empirischen Charakter des Handelnden nicht als etwas, das selbst als gegeben und einer weiteren empirischen Erklärung unzugänglich ist, sondern geben die zeitlich vorhergehenden Ursachen dafür an, dass eine bestimmte Person eben denjenigen Charakter hat, der zu der betreffenden Handlung geführt hat. Diese Kantische Annahme deckt sich mit unserer tatsächlichen Praxis. Wenn aber auch der empirische Charakter durch zeitlich frühere Quellen kausal determiniert ist, dann hätte eine andere transzendental freie Handlung nicht nur Konsequenzen für den empirischen Charakter des Handelnden, sondern auch für das, wodurch dieser Charakter empirisch determiniert ist.

Man kann diesen Punkt wieder in Form eines formal gültigen Arguments darstellen (wieder soll „ $C_E$ “ für den tatsächlichen empirischen Charakter des Lügenden stehen und „ $Z$ “ diesmal für einen Naturzustand, der vor der Ausbildung von  $C_E$  besteht und dafür kausal relevant ist):

- (1+) Es ist unmöglich, dass  $Z$  existiert und  $N_1 \dots N_n$  gelten, aber der Handelnde nicht  $C_E$  hat.
- (2+) Hätte  $H$  nicht stattgefunden, dann hätte der Handelnde nicht  $C_E$  gehabt. ( $= (K_1^*)$ )
- ( $K_0+$ ) Also: Hätte  $H$  nicht stattgefunden, dann hätte entweder  $Z$  nicht existiert oder aber  $N_1 \dots N_n$  nicht gegolten.

Ein ‚altered laws‘-Kompatibilist wäre als Reaktion auf ( $K+$ ) wieder gezwungen die Konsequenz ( $K_1$ ) („Hätte  $H$  nicht stattgefunden, dann hätten  $N_1 \dots N_n$  nicht gegolten“) zu akzeptieren, nur dass man diesmal diese Konsequenz nicht mit dem Hinweis schmackhaft machen kann, dass es sich bei den Verletzungen der Naturgesetze um solche handelt, die sich unmittelbar aus der Änderung des empirischen Charakters des Handelnden ergeben. Schließlich ergibt sich dieser Charakter für Kant – wie eben gesehen – ja aus Dingen wie der Erziehung des Handelnden und seinem gesellschaftlichen Umfeld, und damit aus Situationen, in die nicht nur er selbst, sondern auch andere Substanzen und deren empirische

und intelligible Charaktere involviert ist. Selbst Watkins müsste also zugestehen, dass im Falle des Ausbleibens von  $H$  auch die empirischen Charaktere anderer Substanzen als des Handelnden selbst anders beschaffen sein müssten, als sie es tatsächlich sind. Und diese Konsequenz mag man immer noch für unakzeptabel halten.<sup>27</sup>

Ist die Konsequenz  $(K_1)$  („Hätte  $H$  nicht stattgefunden, dann hätten  $N_1 \dots N_n$  nicht gegolten“) tatsächlich so abwegig? Ich denke, die Antwort muss klarerweise „Nein“ lauten. Schließlich ähnelt  $(K_1)$  einer These, die im Rahmen von heute üblichen Versionen des Kompatibilismus für völlig akzeptabel gehalten wird. Zwar meinen Vertreter dieser Art von Kompatibilismus nicht, dass eine zeitliche Handlung wie die Lüge  $L$  kontrafaktisch von einer zeitlosen und unverursachten Handlung  $H$  abhängen muss, um zurechenbar zu sein. Aber viele von ihnen meinen, dass Freiheit die Fähigkeit, anders zu handeln, impliziert, und sie akzeptieren für  $L$  selbst genau diejenige Konsequenz, die  $(K_1)$  für  $H$  behauptet, nämlich:

$(K_3)$  Hätte  $L$  nicht stattgefunden, dann hätten  $N_1 \dots N_n$  nicht gegolten.

$(K_3)$  ergibt sich aus der Standardtheorie zur Bewertung kontrafaktischer Konditionale, wie man sie etwa bei David Lewis finden kann.<sup>28</sup> Dieser Theorie zufolge ist ein kontrafaktisches Konditional der Form „Wenn  $p$  der Fall gewesen wäre, wäre  $q$  der Fall gewesen“ genau dann wahr, wenn  $q$  in all denjenigen  $p$ -Welten wahr ist, die der wirklichen Welt am ähnlichsten sind.<sup>29</sup> Laut Lewis' Kriterien für Weltenähnlichkeit sind solche Nicht- $L$ -Welten der wirklichen Welt am ähnlichsten, die bis kurz vor  $L$  der wirklichen Welt vollständig gleichen und in der dann kurz vor  $L$  eine Verletzung der in der wirklichen Welt geltenden Naturgesetze (ein „kleines Wunder“) passiert, die das Ausbleiben von  $L$  bewirkt, wobei nach dieser Verletzung wieder alles so geschieht, wie die tatsächlichen Naturgesetze es vorschreiben. Lewis' Theorie hat zur Folge, dass  $(K_3)$  wahr ist.

In seinem Aufsatz „Are we free to break the laws?“<sup>30</sup> zeigt Lewis ferner, dass aus der Wahrheit von  $(K_3)$  und der Annahme, dass der Lügner die Fähigkeit hatte, die Lüge zu

---

<sup>27</sup> Watkins setzt sich mit diesem Einwand kurz in einer Fußnote auseinander und schreibt: „One might object that our empirical natures are caused by prior events so that our choice of our empirical natures really does entail prior events. However, there is an important distinction between causing a nature that is instantiated in the world to be efficacious in certain ways and causing a nature to be instantiated in the world“ (*Watkins: Metaphysics of Causality* (Anm. 24), 336, Fußnote 38). Mir ist nicht klar, auf welche Weise genau die genannte Unterscheidung helfen soll, das Problem zu vermeiden. Schließlich scheint sowohl die Existenz eines bestimmten empirischen Charakters in der Welt als auch die Aktivierung eines in der Welt existierenden Charakters durch zeitlich frühere Zustände determiniert zu sein.

<sup>28</sup> Vgl. *David Lewis: Counterfactual Dependence and Time's Arrow*. – In: *Nous* 13 (1979). 455–76.

<sup>29</sup> Dies ist die vereinfachte Version von Lewis' eigentlicher Definition, welche auf die Voraussetzung verzichten kann, dass es Welten gibt, die unserer Welt „näher“ als alle anderen Welten sind. Da diese Voraussetzung in unserem Kontext unwichtig ist und die vereinfachte Version leichter fasslich ist, werde ich weiter mit ihr arbeiten.

<sup>30</sup> In: *Theoria*. 47 (1981). 113–121.

unterlassen, nicht folgt, dass wir dem Lügner so etwas wie die Fähigkeit zusprechen müssen, die (tatsächlichen) Naturgesetze außer Kraft zu setzen. Dies wäre tatsächlich eine unakzeptable Konsequenz, aber es wäre auch nur dann der Fall, wenn Lewis' Theorie die Wahrheit des folgenden kontrafaktischen Konditionals implizieren würde:

(K<sub>4</sub>) Hätte *L* nicht stattgefunden, dann hätte das Unterlassen von  $L N_1 \dots N_n$  verletzt.

In anderen Worten: Die Naturgesetze würden wir nur dann außer Kraft setzen, wenn wir etwas täten, das entweder selbst den Naturgesetzen widerspricht oder ein naturgesetzverletzendes Ereignis verursacht. Laut Lewis' Analyse ist es aber nicht die Unterlassung der Lüge selbst, die eine Verletzung der Naturgesetze darstellt, sondern das dieser Unterlassung vorhergehende sog. „kleine Wunder“. (Wie Lewis deutlich macht, ist dieses Wunder auch nicht die Wirkung der Unterlassung.)

Es ist natürlich abwegig, Kant ein wenn auch nur implizites Verständnis von Lewis' Analyse kontrafaktischer Konditionale zu unterstellen. Aber darum geht es hier nicht. Es geht um die Frage, ob eine Konsequenz wie (K<sub>1</sub>) systematisch so absurd ist, dass man eine Interpretation ablehnen muss, bei der sich diese Konsequenz aus Kants Kompatibilismus ergibt. Und diese Frage kann man mit Verweis auf Lewis klar verneinen. Ob die vorgeschlagene Verteidigung im Rahmen von Kants Philosophie akzeptabel ist, hängt allerdings natürlich auch davon ab, welche Auffassung über den modalen Status von Naturgesetzen man Kant zuschreibt, denn die Verteidigung funktioniert nur, wenn es mit dem modalen Status der in (K<sub>1</sub>) und (K<sub>3</sub>) genannten Naturgesetze  $N_1 \dots N_n$  vereinbar ist, dass sie in einer unserer Welt sehr ähnlichen Welt nicht gelten.<sup>31</sup> Eine Auseinandersetzung mit dieser Problematik würde eine Beschäftigung mit der umfangreichen und kontroversen Debatte zum Status spezieller Naturgesetze bei Kant erfordern und sprengt deswegen den Rahmen dieses Beitrags. Die Tatsache, dass eine Verteidigung des ‚altered laws‘-Kompatibilismus an diesem Punkt unter Umständen mit Kantischen Annahmen konfligiert, lässt es aber lohnend erscheinen, zu untersuchen, ob man Kants Position auch als ‚altered past‘-Kompatibilismusinterpretieren könnte, ohne ihm abwegige Ansichten zuschreiben zu müssen.

#### 4. *Are we free to change the past?*

Sehen wir uns noch einmal das Argument an, aus dem sich die Schwierigkeit für Watkins' Ansatz ergeben hatte:

---

<sup>31</sup> Laut Lewis' sog. „best system analysis“-Konzeption von Naturgesetzen ist diese Annahme kein Problem.

- (1+) Es ist unmöglich, dass  $Z$  existiert und  $N_1 \dots N_n$  gelten, aber der Handelnde nicht  $C_E$  hat.
- (2+) Hätte  $H$  nicht stattgefunden, dann hätte der Handelnde nicht  $C_E$  gehabt.
- ( $K_0+$ ) Also: Hätte  $H$  nicht stattgefunden, dann hätte entweder  $Z$  nicht existiert oder aber  $N_1 \dots N_n$  nicht gegolten.

Wer nicht akzeptieren will, dass im Falle der Unterlassung von  $H$  die Naturgesetze andere gewesen wären, oder zumindest nicht, dass sie über die allein aus der Änderung des empirischen Charakters  $C_E$  resultierende Weise hinaus andere gewesen wären, ist gezwungen, die folgende Konsequenz zu akzeptieren:

- ( $K_2$ ) Hätte  $H$  nicht stattgefunden, dann hätte  $Z$  nicht existiert.

Aus dem Unterlassen von  $H$  folgt also, dass ein zeitlich vor der Lüge vorhergehender Naturzustand nicht existiert hätte. Da dieser Zustand  $Z$  seinerseits kausal durch die Naturgesetze und durch einen zeitlich vor  $Z$  liegenden Zustand  $Z'$  determiniert ist, hätte das Unterlassen von  $H$  auch das Ausbleiben von  $Z'$  zur Folge, und das Ausbleiben eines  $Z'$  determinierenden Zustands  $Z''$ , und immer so weiter. Wir können diese Konsequenz folgendermaßen formulieren:

- ( $K_2^*$ ) Hätte  $H$  nicht stattgefunden, dann hätten beliebig weit in die Vergangenheit zurückreichende Abfolgen von Zuständen  $Z, Z', Z'' \dots Z^n$ , die jeweils darin resultieren, dass der Handelnde  $C_E$  hat, nicht stattgefunden.

Die Form von ‚altered past‘-Kompatibilismus<sup>32</sup>, die in der Akzeptanz von ( $K_2^*$ ) besteht, hat den exegetischen Vorteil, dass Kant sie an einer Stelle explizit vertritt. In der *Kritik der praktischen Vernunft* schreibt er:

„[...] die Bestimmungsgründe einer jeden Handlung [des handelnden Subjects als Erscheinung; T.R.] liegen [in demjenigen], was zur vergangenen Zeit gehört und nicht mehr in seiner Gewalt ist (wozu auch seine schon begangene Thaten und der ihm dadurch bestimmbare Charakter in seinen eigenen Augen, als Phänomens, gezählt werden müssen). Aber ebendasselbe Subject, das sich andererseits auch seiner als Dinges an sich selbst bewußt ist, betrachtet auch sein Dasein, so fern es nicht unter Zeitbedingungen steht, [...] und in diesem seinem Dasein ist ihm nichts vorhergehend vor seiner Willensbestimmung, sondern jede Handlung [...], selbst die ganze Reihenfolge seiner Existenz als Sinnenwesen ist im

<sup>32</sup> Am prominentesten vertreten von Allen Wood in *Kant's Compatibilism*. – In: Allen Wood (Ed.): *Self and Nature in Kant's Philosophy*. Ithaca 1984. 73–101.

Bewußtsein seiner intelligibelen Existenz nichts als Folge, niemals aber als Bestimmungsgrund seiner Causalität, als Noumens, anzusehen. In diesem Betracht nun kann das vernünftige Wesen von einer jeden gesetzwidrigen Handlung, die es verübt, ob sie gleich als Erscheinung in dem Vergangenen hinreichend bestimmt und so fern unausbleiblich notwendig ist, mit Recht sagen, daß er sie hätte unterlassen können; *denn sie mit allem Vergangenen, das sie bestimmt, gehört zu einem einzigen Phänomen seines Charakters*, den er sich selbst verschafft, und nach welchem er sich als einer von aller Sinnlichkeit unabhängigen Ursache die Causalität jener Erscheinungen selbst zurechnet.“ (AA V 98 f.; Hervorhebung von mir; T.R.)

Mit dem Charakter, den sich das handelnde Subjekt selbst verschafft, ist, wie oben erläutert, der intelligible Charakter gemeint. Kants Aussage, dass „alles Vergangene“, das eine zeitliche Handlung determiniert, „zum Phänomen“ dieses Charakters gehört, kann man dann im Sinne der Konsequenz ( $K_2^*$ ) verstehen: Hätte sich das Subjekt mit seiner transzendental freien Handlung einen anderen intelligiblen Charakter verschafft und wäre die gesetzeswidrige Handlung deswegen ausgeblieben, dann wäre auch die gesamte kausale Vorgeschichte der Handlung eine andere gewesen.

Ist ( $K_2^*$ ) eine akzeptable Konsequenz? Derk Pereboom bezeichnet einen ‚altered past‘-Kompatibilismus, der ( $K_2^*$ ) akzeptiert, als „at best insignificantly more credible than an overt contradiction“<sup>33</sup>. Das sind starke Worte – was könnte sie rechtfertigen? Ich denke, es sind zwei Einwände, die man gegen ( $K_2^*$ ) vorbringen könnte – und die zumindest teilweise auch explizit dagegen artikuliert worden sind. Der erste Einwand lautet, dass aus ( $K_2^*$ ) folgt, dass  $H$  eine Ursache für den bisherigen Weltverlauf ist und unsere Fähigkeit, transzendental frei zu handeln, also die Fähigkeit implizieren würde, kausal auf die Vergangenheit einzuwirken. Der zweite Einwand lautet, dass ( $K_2^*$ ) impliziert, dass wir nicht nur für die gesetzeswidrige Handlung, die durch ein Unterlassen von  $H$  zu verhindern gewesen wäre, moralisch verantwortlich sind, sondern zudem für die gesamte kausale Vorgeschichte dieser Handlung. Ich werde im folgenden beide Einwände darstellen und dann zu entkräften versuchen.

(i) Impliziert ( $K_2^*$ ), dass  $H$  eine Ursache für den bisherigen Weltverlauf ist, so dass die Fähigkeit transzendental frei zu handeln die Fähigkeit impliziert, kausal auf die Vergangenheit einzuwirken? Ein sachlicher Grund für diese Annahme könnte darin bestehen, das Bestehen einer kontrafaktischen Abhängigkeit der Form „Wenn  $p$  nicht der Fall gewesen wäre, wäre auch  $q$  nicht der Fall gewesen“ als hinreichend dafür anzusehen, dass  $p$  eine Ursache dafür ist, dass  $q$ . Wenn man dies annimmt, impliziert ( $K_2^*$ ), dass die ~~transzendental freie Handlung~~  $H$  eine Ursache für die Existenz der Naturzustände  $Z, \dots, Z^n$

<sup>33</sup> Pereboom: *Kant on Transcendental Freedom* (Anm. 4). 556.

Handlung  $H$  eine Ursache für die Existenz der Naturzustände  $Z, \dots, Z''$  ist. Diese Konsequenz scheint deswegen absurd, weil es sich bei  $Z, \dots, Z''$  um bereits vergangene Naturzustände handelt und sich die Vergangenheit dadurch auszeichnet, dass man nicht kausal auf sie Einfluss nehmen kann. Es ist allerdings zu beachten, dass man im Falle von  $H$  gar nicht davon sprechen kann, dass zum Zeitpunkt von  $H$  die Zustände  $Z, \dots, Z''$  bereits vergangen sind und  $H$  also eine kausale Wirkung in die Vergangenheit hinein entfaltet.  $H$  ist schließlich gar nicht eine Handlung, die zu einem bestimmten Zeitpunkt stattfindet. Absurd wäre zweifellos das folgende Bild:  $Z, \dots, Z''$  sind bereits vergangen, dann kommt es zu der transzendental freien Handlung  $H$ , und nun soll gelten, dass  $Z, \dots, Z''$  bestehen, *weil* es zu  $H$  gekommen ist und  $H$  also  $Z, \dots, Z''$  im Nachhinein verursachen. Aber so stellt sich Kant das ja auch nicht vor.  $H$  ist eine Handlung des Menschen als Ding an sich und als solches existiert er nicht in der Zeit, so dass auch  $H$  nicht zum Zeitpunkt z.B. der Lüge oder kurz vorher stattfindet.

Der Hinweis auf die Zeitlosigkeit von  $H$  allein reicht allerdings nicht aus, um die genannte Schwierigkeit wirklich auszuräumen. Selbst wenn  $H$  nicht in der Zeit stattfindet und also ein kausaler Einfluss auf  $Z, \dots, Z''$  nicht mit einem kausalen Einfluss auf die Vergangenheit gleichzusetzen ist, findet die zeitliche Handlung, deren moralische Zurechenbarkeit  $H$  garantieren soll, – z.B. die Lüge  $L$  – doch in der Zeit statt.  $(K_2^*)$  ist eigentlich nur eine Folge der Tatsache, dass  $L$  nicht stattgefunden hätte, wenn  $H$  nicht stattgefunden hätte, und dass das folgende kontrafaktische Konditional wahr ist:

- $(K_3)$  Hätte  $L$  nicht stattgefunden, dann hätten beliebig weit in die Vergangenheit zurückreichende Folgen von Zuständen  $Z, \dots, Z''$ , die jeweils in  $L$  resultieren, nicht stattgefunden.

Wenn kontrafaktische Abhängigkeit hinreichend für kausalen Einfluss ist, hätte also auch  $L$  einen kausalen Einfluss auf  $Z, \dots, Z''$ , wobei dieser kausale Einfluss nun unabweislich einer auf die Vergangenheit wäre, schließlich handelt es sich bei  $L$  ja um eine Handlung in der Zeit.

Um diesem Vorwurf zu begegnen, ist es wieder hilfreich, einen kurzen Blick in die gegenwärtige Diskussion zur Bewertung kontrafaktischer Konditionale zu werfen. Kant selbst behauptet ja nirgends, dass die Lüge eine Ursache für ihre eigene Vorgeschichte ist; also wäre der einzige Grund, ihm diese absurde Annahme als Konsequenz seines Kompatibilismus zuzuschreiben, der, dass man  $(K_3)$  nicht vertreten kann, ohne aus systematischen Gründen zu dieser Annahme gezwungen zu sein. Gemäß der Standardtheorie über die Wahrheitsbedingungen kontrafaktischer Konditionale ist ein kontrafaktisches Konditional wie gesagt genau dann wahr, wenn das Konsequens in allen der wirklichen Welt ähnlichsten Welten wahr ist, in denen das Antezedens wahr ist. Nun sind die allermeisten Theore-

tiker, die diese Standardauffassung akzeptieren, der Meinung, dass Welten, in denen kleine Verletzungen der Naturgesetze geschehen, der wirklichen Welt ähnlicher sind, als solche, in denen genau die selben Naturgesetze gelten, wie in der wirklichen Welt, dafür aber der Weltverlauf vor der Antezedens-Situation massiv von der tatsächlichen Vergangenheit abweicht. Jonathan Bennett hat allerdings zumindest eine Zeit lang eine Position vertreten, der zufolge in den unserer Welt ähnlichsten Welten ausnahmslos unsere Naturgesetze gelten und die Abweichung, die zum Bestehen der Antezedenssituation führt, dadurch zustande kommt, dass sich der Weltverlauf bis zu dieser Situation von der tatsächlichen Welt unterscheidet.<sup>34</sup> Ein kontrafaktisches Konditional der Form „Wenn  $p$  der Fall (gewesen) wäre, wäre auch  $q$  der Fall (gewesen)“ ist Bennetts Vorschlag zufolge genau dann wahr, wenn  $q$  in allen  $p$ -Welten wahr ist, die der tatsächlichen Welt zum Zeitpunkt des Bestehens der  $p$ -Situation am nächsten sind und in der unsere Naturgesetze gelten.<sup>35</sup> Bennett hat diese Position entwickelt, um der Tatsache gerecht zu werden, dass bestimmte sog. „backtracking counterfactuals“, also kontrafaktische Konditionale, bei denen das Konsequens von einer Situation handelt, der derjenigen, von der im Antezedens die Rede ist, zeitlich vorhergeht, wahr zu sein scheinen.<sup>36</sup>

Für unseren Zusammenhang ist wichtig, dass eine solche Theorie der Wahrheitsbedingungen kontrafaktischer Konditionale auf jeden Fall die Wahrheit von  $(K_3)$  impliziert. Nun hat Bennett zwar inzwischen die genannte Theorie wieder verworfen. Aber der Grund dafür ist, dass diese Theorie für bestimmte Fälle von „backtracking counterfactuals“ unintuitive Ergebnisse liefert; der Grund ist nicht, dass diese Theorie und  $(K_3)$  von vornherein absurd und „at best insignificantly more credible than an overt contradiction“ ist, um noch einmal an die Formulierung von Perebooms Kritik zu erinnern. Vor allem aber liegt der Grund für die Ablehnung sicher nicht darin, dass man der Theorie die Implikation zuschreiben muss, dass Ereignisse Ursachen für zeitlich frühere Ereignisse sein können. Es ist nämlich völlig klar – und wird von Bennett eher nebenbei festgestellt<sup>37</sup> –, dass eine Theorie, die ‚backtracking counterfactuals‘ erlaubt, nicht zulassen darf, dass diese kontrafaktischen Konditionale in einer kontrafaktischen Analyse der Kausalrelation eine Rolle spielen, wenn man diese Analyse denn überhaupt plausibel findet. Diese Replik auf den Einwand stünde Kant selbstverständlich auch offen und es ist deswegen keineswegs so, dass er aus systematischen Gründen durch die Akzeptanz von  $(K_3)$  zu der Annahme gezwungen ist, dass die Lüge eine Ursache für ihre eigene kausale Vorgeschichte ist.

---

<sup>34</sup> Vgl. Jonathan Bennett: *Counterfactuals and Temporal Direction* – In: *The Philosophical Review*. 93 (1984). 57-91.

<sup>35</sup> Bennett lässt offen, worin die Nähe einer Welt zur wirklichen Welt zu einem Zeitpunkt besteht und legt nur fest, dass sie auf jeden Fall Ähnlichkeit zu diesem Zeitpunkt impliziert.

<sup>36</sup> Ein Beispiel wäre etwa: „Wenn Steinmeier heute Kanzler wäre, wäre er 2009 gewählt worden“.

<sup>37</sup> Bennett: *Counterfactuals and Temporal Direction*. 68.

(ii) Der zweite Einwand gegen den ‚altered past‘-Kompatibilismus geht auf Ralph Walker zurück.<sup>38</sup> Walker meint, dass Kants Kompatibilismus die unangenehme Konsequenz hat, dass Menschen nicht nur für ihre in der Zeit stattfindenden Handlungen moralisch verantwortlich sind, sondern zudem auch für die gesamte kausale Vorgeschichte, die zu diesen Handlungen führt. Walkers Grund für diese Annahme besteht offensichtlich darin, dass Kant durch  $(K_2^*)$  gezwungen ist, für beliebige Zustände  $Z$ , die zeitlich vor der Ausbildung des empirischen Charakters eines Menschen liegen und Teil der kausalen Vorgeschichte dieses Charakters sind, anzunehmen, dass  $Z$  nicht existiert hätte, wenn die zeitlose Aneignung  $H$  des Charakters nicht stattgefunden hätte, und dass der Mensch deswegen moralisch für  $Z$  verantwortlich ist. Mit drastischen Beispielen macht Walker klar, wie absurd diese Konsequenz ist: „I can be blamed for the First World War and for the Lisbon earthquake that so appalled Voltaire. Gandhi is no less guilty than Amin of the atrocities of the Ugandan dictator.“<sup>39</sup>

Allen Wood, der vielleicht prominenteste Verfechter des ‚altered past‘-Kompatibilismus, hat auf Walkers Einwand mit der Behauptung reagiert, dass man einen Menschen nicht für *jeden* Zustand, der Teil der kausalen Vorgeschichte seines empirischen Charakters ist, verantwortlich machen muss, sondern nur für solche, die zu einer kausalen Vorgeschichte dieses Charakters gehören *müssen*.<sup>40</sup> Wood erläutert das theoretische Fundament dieser Einschränkung nicht weiter und führt sie einfach als intuitiv plausibel ein, aber man kann sie vielleicht anhand von Bennetts eben skizzierter Theorie begründen: Durch die zeitlose Aneignung  $H$  des intelligiblen Charakters soll festgelegt werden, ob der Handelnde ab einem Zeitpunkt  $t$  den empirischen Charakter  $C_E$  hat oder einen anderen empirischen Charakter. Betrachten wir nun ein beliebiges kontrafaktisches Konditional der Form „Wenn der Handelnde zu  $t$   $C_E$  hätte, dann  $q$ “. Um wahr zu sein, muss  $q$  laut den Bennett’schen Wahrheitsbedingungen für kontrafaktische Konditionale in *allen* möglichen Welten wahr sein, in denen der Handelnde zu  $t$   $C_E$  hat und die tatsächlichen Naturgesetze gelten und die der wirklichen Welt zu  $t$  am nächsten sind. D.h. es ist nur dann wahr, wenn  $q$  in *jeder* der relevanten kausalen Vorgeschichten, die dazu führen, dass der Handelnde  $C_E$  hat, vorkommt.  $q$  scheint also genau von denjenigen Zustände zu handeln, von denen auch Wood in seiner Einschränkung spricht. Kants Position kann laut Wood also dahingehend verstanden werden, dass ein Handelnder für jeden Sachverhalt  $q$  moralisch verantwortlich ist, für den gilt, dass er bestehen würde, wenn der Handelnde zu  $t$   $C_E$  hätte. Jonathan Bennett hat – diesmal in seiner Funktion als Kantforscher – gegen Woods Vorschlag eingewandt, dass dieser Vorschlag immer noch zu absurden Konsequenzen führt. Er schreibt: „One might think, for example, that in 1929 oxygen exists not in all deterministic worlds but in

---

<sup>38</sup> *Ralph Walker: Kant*. London, Henley, Boston 1978. 149.

<sup>39</sup> Ebd.

<sup>40</sup> *Wood: Kant’s Compatibilism* (Anm. 32). 92.

all the ones where I am born in 1930 with character E; and so by Kant's theory I am morally responsible for the presence of oxygen in the universe in 1929."<sup>41</sup>

Wenn Kants Kompatibilismus beinhalten würde, dass jemand heute für den Ersten Weltkrieg oder für die Anwesenheit von Sauerstoff im Jahr 1929 moralisch verantwortlich ist, käme dies in der Tat einer Widerlegung gleich. Aber ist Kants Kompatibilismus wirklich auf diese Konsequenz festgelegt? Zwischen Walker, Wood und Bennett besteht ein erstaunliches Einvernehmen darüber, dass man aus einer bestimmten Form der kontrafaktischen Abhängigkeit eines Zustands *Z* von einer transzendental freien Handlung *H* darauf schließen darf, dass der Handelnde für das Bestehen von *Z* moralisch verantwortlich ist. Bei Walker ist das die Abhängigkeit, die zwischen *H* und *Z* besteht, weil *Z* nicht bestanden hätte, wenn *H* nicht stattgefunden hätte. Bei Wood ist es die Abhängigkeit, die zwischen *H* und *Z* besteht, weil *Z* auf jeden Fall bestehen müsste, wenn *H* stattfinden würde. Die drei Autoren schweigen sich jedoch leider darüber aus, weshalb sie der Meinung sind, dass dieser Zusammenhang zwischen kontrafaktischer Abhängigkeit und moralischer Verantwortung besteht, und man fragt sich, weshalb um alles in der Welt er das tun sollte. Das einzige Argument, das einem in den Sinn kommt, geht davon aus, dass die kontrafaktische Abhängigkeit der Lüge *L* von *H* bei Kant erklären soll, weshalb wir für *L* moralisch verantwortlich sein können, und schließt dann, dass die selbe kontrafaktische Abhängigkeit auch zwischen *Z* und *H* besteht. Macht man das in dieser Überlegung enthaltene Argument explizit, erhält man – im Falle von Walkers Version – den folgenden Schluss:

- (i) Der Handelnde ist nur dann moralisch für *L* verantwortlich, wenn *L* nicht stattgefunden hätte, wenn die transzendental freie Handlung *H* nicht stattgefunden hätte.
- (ii) *Z* hätte nicht stattgefunden, wenn die transzendental freie Handlung *H* nicht stattgefunden hätte.
- (iii) Also ist der Handelnde moralisch für *Z* verantwortlich.

So formuliert sieht man sofort, dass Walkers Überlegung ein recht drastisches *non sequitur* zu Grunde liegt. Nur weil die kontrafaktische Abhängigkeit der Lüge von *H* eine *notwendige Bedingung* dafür ist, dass der Handelnde für *L* verantwortlich ist, sind analoge kontrafaktische Abhängigkeiten anderer Zustände von *H* noch lange nicht *hinreichend* dafür, dass der Handelnde für diese Zustände verantwortlich ist.

Sachlich gesehen ist die Annahme, dass man für alle Zustände moralisch verantwortlich ist, die nicht bestanden hätten oder nicht bestehen würden, wenn man sich anders entschieden hätte – so etwa für alle Wirkungen des eigenen Handelns –, natürlich eine ausgewachsene Absurdität. Deswegen sollte man diese Annahme Kant nur dann zuschreiben, wenn seine

---

<sup>41</sup> Jonathan Bennett: *Kant's Theory of Freedom*. – In: Allen Wood (Ed.): *Self and Nature in Kant's Philosophy*. Ithaca 1984. 102–112. 103.

eigenen Schriften starke Hinweise darauf enthalten, dass er sie akzeptiert hat.<sup>42</sup> Ansonsten ist es naheliegend, Kant die Ansicht zu unterstellen, dass die Verantwortlichkeit für eine Lüge nicht allein durch die kontrafaktische Abhängigkeit der Lüge von  $H$  bedingt ist, sondern dass die Lüge darüber hinaus auch noch andere Bedingungen erfüllen muss, um verantwortlich zu sein, z.B. das unmittelbare Ergebnis des Subsumierens der Handlungssituation unter eine der eigenen Maximen sein muss. Diese Bedingung würden der erste Weltkrieg und die Anwesenheit von Sauerstoff in der Welt nicht erfüllen.

### *5. Kontrafaktische Abhängigkeit und wirkende Ursachen*

Bislang wurde Kants Reaktion auf das im Anfangszitat beschriebene Dilemma folgendermaßen rekonstruiert: Die erwähnte „boshafte Lüge“ ist kausal prädeterminiert, weil ihre Existenz aus den Handlungsumständen und aus dem empirischen Charakter des Lügners folgt. Zudem ist auch der empirische Charakter des Lügners vollständig durch den Weltverlauf vor der Entstehung dieses Charakters festgelegt. Dass wir den Lügner trotzdem für die Lüge verantwortlich machen, liegt daran, dass wir ihn für seinen empirischen Charakter verantwortlich machen, und zwar deswegen, weil ein bestimmtes Verhältnis zwischen dem empirischen Charakter, dem intelligiblen Charakter und dem transzendental freien, d.h. unverursachten, und außerzeitlichen intelligiblen Handeln des Lügners als eines Dinges an sich besteht. Dieses Verhältnis wurde bislang als das der kontrafaktischen Abhängigkeit spezifiziert: Hätte die besagte intelligible Handlung nicht stattgefunden, d.h. hätte sich der Lügner nicht einen intelligiblen Charakter verschafft, der als oberste Maxime die Unterordnung der sittlichen unter die sinnliche Triebfeder beinhaltet, dann hätte der Lügner auch nicht denjenigen empirischen Charakter gehabt, der zusammen mit den Handlungsumständen zu der Lüge führt. Deswegen ist er für seine Lüge verantwortlich. Wie in den letzten beiden Abschnitten dargelegt, kann Kant die genannte kontrafaktische Abhängigkeit behaupten, ohne sich dadurch allzu abwegige Konsequenzen einzuhandeln.

Ich möchte abschließend nun einen Einwand diskutieren, den man gegen die bislang vorgestellte Rekonstruktion und Verteidigung von Kants Kompatibilismus vorbringen könnte. Er lässt sich in zwei Schritte untergliedern und lautet folgendermaßen:

- (i) Es mag zwar einleuchtend sein, dass Kant die moralische Zurechnungsfähigkeit einer Lüge  $L$  dadurch erklärt, dass der empirische Charakter  $C_E$  des Lügners, aus dem zusammen mit den Handlungsumständen die Lüge folgt, ein anderer gewesen wäre, wenn der Lügner sich nicht durch eine bestimmte transzendental freie, intelligible Handlung

---

<sup>42</sup> Und selbst wenn dies der Fall wäre, würde sich daraus kein Argument gegen Kants Kompatibilismus ergeben, sondern Kant wäre vielmehr für eine absurde Auffassung über moralische Zurechenbarkeit zu kritisieren.

$H$  einen bestimmten intelligiblen Charakter  $C_I$  zugelegt hätte. Aber das von Kant angenommene Verhältnis zwischen  $H$  einerseits und  $C_E$  bzw.  $L$  andererseits kann sich nicht in dieser kontrafaktischen Abhängigkeit erschöpfen. Diese wäre im Rahmen von Kants Verständnis von Kausalität nämlich nicht hinreichend dafür, die Handlung  $H$  eine *Ursache* von  $L$  oder von  $C_E$  nennen zu können – was Kant aber tut.<sup>43</sup> Für Kant ist etwas nämlich nicht schon dadurch eine Ursache, dass seine Wirkung nicht auftreten würde, wenn es selbst nicht auftreten würde, sondern erst dadurch, dass es selbst nicht auftreten kann, ohne dass seine Wirkung auftritt: „Der Begriff der Ursache enthält eine Regel, nach der aus einem Zustande ein anderer nothwendiger Weise folgt“ (AA IV 314). Auf den vorliegenden Fall angewandt heißt das: Die vorgestellte Verteidigung von Kants Kompatibilismus nimmt an, dass  $H$  eine *notwendige* Bedingung dafür ist, dass  $L$  auftritt und der Handelnde  $C_E$  hat. Kant behauptet aber, dass  $H$  eine *hinreichende* Bedingung dafür ist, dass  $L$  auftritt und der Handelnde  $C_E$  hat.

- (ii) Versteht man  $H$  als hinreichende Bedingung für  $L$  und  $C_E$ , ergeben sich aber wieder genau diejenigen Probleme, die die in Abschnitt 3 und 4 vorgestellte Verteidigung von Kants Kompatibilismus auszuräumen versucht hat. Dass  $H$  eine hinreichende Bedingung für  $L$  und  $C_E$  ist, kann man als die These verstehen, dass eine mögliche Welt, in der  $H$  stattfindet, immer auch eine ist, in der  $L$  stattfindet und der Handelnde  $C_E$  hat. In diesem Fall sorgt der Handelnde allein dadurch, dass er  $H$  ausführt, dafür, dass die wirkliche Welt eine ist, in der  $L$  stattfindet und er selbst  $C_E$  hat. Dann aber scheint es plausibel anzunehmen, dass der Handelnde, allein indem er  $H$  ausführt, dafür sorgt, dass die wirkliche Welt eine ist, deren Weltverlauf vor  $L$  zusammen mit den in ihr geltenden Naturgesetze dazu führt, dass  $L$  auftritt. Das würde zwar – ganz im Sinne des Vorschlags von Wood – nicht heißen, dass durch unsere freie Handlung genau der tatsächliche Weltverlauf vor  $L$  und die tatsächlich geltenden Naturgesetze als diejenigen in der wirklichen Welt festgelegt werden, denn  $H$  legt ja nur fest, dass unsere Welt *irgendwie* so beschaffen sein muss, dass  $L$  auftritt. Aber man kann immer noch einwenden, dass es sehr unplausibel ist, dass sich die wirkliche Welt von einem frei handelnden Menschen überhaupt irgendwelche Vorschriften hinsichtlich der Frage machen lassen sollte, welche Naturgesetze in ihr gelten oder welchen Verlauf sie vor der zeitlichen Existenz des betreffenden Menschen genommen hat.<sup>44</sup>

<sup>43</sup>  $H$  ist zweifellos als Ursache des intelligiblen Charakters zu verstehen; zudem behauptet Kant von diesem intelligiblen Charakter sowohl, dass er die „transscendentale Ursache“ des empirischen Charakters sei (A 546/B 574), als auch, dass jede empirische Handlung seine „unmittelbare Wirkung“ sei (KrV A 553/B 581).

<sup>44</sup> Wolfgang Ertl hat in verschiedenen Schriften zu Kants Freiheitslehre dafür argumentiert, dass Kant über die Ressourcen verfügt, der eben genannten Konsequenz ihren philosophischen Schrecken zu nehmen. Ertl vertritt eine kosmologisch-theologische Version des ‚altered laws‘-Kompatibilismus, der zufolge es Gott ist, der unsere freien intelligiblen Handlungen zur Kenntnis nimmt und dann die Menge der speziellen Naturgesetze so bestimmt, dass die deterministische Welt mit dem phänomenalen Ergebnis dieser Handlungen vereinbar ist (vgl. z.B. *Ertl: Schöpfung und Freiheit* (Anm. 24)). Ertl zeigt ferner, dass Kant sich die Inspiration für diese Konzeption wiederum bei Luis de Molina verschafft haben könnte (vgl. *Ertl: Molina and Kant's Liberta-*

Meine Replik auf diesen Einwand wird in dem Nachweis bestehen, dass Kant Ursachen im allgemeinen und die transzendental freie Handlung *H* im besonderen nicht als hinreichende Bedingungen in dem für den Einwand maßgeblichen Sinn versteht. In den *Prolegomena* wählt Kant als paradigmatischen Fall eines Kausalverhältnisses das Warmwerden eines Steins durch das Scheinen der Sonne (vgl. AA IV 301 Fn. und 311). Ist das Scheinen der Sonne in dem für den Einwand vorausgesetzten Sinne allein hinreichend dafür, dass sich der Stein erwärmt? Sicher nicht, schließlich gibt es viele mögliche Welten, in denen die Sonne scheint, ohne dass sich der Stein erwärmt, zum Beispiel Welten, in denen der Stein gar nicht existiert, sich nicht in der Nähe der Sonne befindet, oder im Schatten einer Mauer liegt.

In der *Metaphysikvorlesung L<sub>2</sub> nach Pölitz* führt Kant eine Terminologie ein, die hilfreich ist, um diejenigen Faktoren zu beschreiben, die vorhanden sein müssen, damit eine Ursache wie das Sonnenscheinen eine Wirkung wie die Steinerwärmung haben kann. Er definiert dort eine Ursache allgemein als einen „Grund der Wirklichkeit“ (AA XXVIII 571) und meint, dass wir zwei Arten solcher Gründe der Wirklichkeit unterscheiden sollten: „Causa efficiens ist eine Ursache durch einwirkende Kraft. Die *Conditio sine qua non* ist eine Bestimmung der Dinge, die zwar nicht negativ ist, aber auch nicht wirkende Ursache heißt, ob sie gleich zur Ursache gerechnet wird. So ist bei der Kanonenkugel das Pulver *conditio sine qua non*; causa efficiens aber der Soldat, der die Kanone abbrennt“ (AA XXVIII 572)<sup>45</sup>. In seinen *Vorlesungen über die philosophische Religionslehre (nach Pölitz)*, schreibt Kant über das Verhältnis verschiedener Ursachen, die nur zusammen eine Wirkung hervorbringen: „Es können nämlich mehrere Ursachen sich vereinigen, um eine Wirkung hervorzubringen. Geschiehet das, so concurriren in solchem Falle mehrere *concausae*. Von diesen mitwirkenden Ursachen muß eine für sich selbst zur Hervorbringung der Wirkung nicht hinreichend seyn; denn sonst wäre die Vereinigung einer andern unnöthig, die ihr das *complementum ad sufficientiam* geben soll“ (AA XXVIII 1105; vgl. auch *Metaphysik Mrogovius* AA XXIX 844 f.). Auf den Fall des Kanonenschusses angewandt heißt das, dass nur der Soldat, das Pulver und wahrscheinlich noch viele weitere Faktoren zusammengenommen hinreichend für den Schuss sind, jeder dieser Faktoren einzeln aber – auch die „wirkende Ursache“, d.h.

---

*rian Compatibilism* (Anm. 23)). Auch Derk Pereboom meint, dass man sich zur Rettung von Kants Kompatibilismus bei Molinas Konzeption göttlicher Vorhersehung freier menschlicher Handlungen bedienen müsse, allerdings ohne wie Ertl die Möglichkeit eines tatsächlichen historischen Bezugs zwischen Kant und Molina nachzuweisen (vgl. *Pereboom: Kant on Transcendental Freedom* (Anm. 4). 557). Sollte sich der Einwand im Haupttext nicht entkräften lassen und Kant transzendental freie Handlungen tatsächlich als hinreichende Bedingungen für empirische Handlungen und Charaktere verstehen, scheint mir die von Ertl und Pereboom vorgeschlagene Konzeption die eleganteste Lösung der Schwierigkeiten mit Kants Kompatibilismus zu sein. Allerdings bringt die These, dass wir laut Kant nicht nur den transzendentalen Idealismus, sondern zudem auch die Existenz Gottes benötigen, um die Vereinbarkeit von Freiheit und Naturdeterminismus zu garantieren, exegetisch sicher einige Härten mit sich.

<sup>45</sup> Dass nicht nur die wirkende Ursache eine Ursache ist, wird auch in der Reflexion 3618 deutlich, in der Kant Baumgartens Definition „CAUSA realitatis per actionem est EFFICIENS“ in § 319 seiner *Metaphysica* (Halle, Magdeburg 1757) durch den Zusatz „*vel efficiens vel conditio sine qua non*“ korrigiert (AA XVII 96).

die *causa efficiens* – nur eine notwendige Bedingung dafür ist, dass die Kanone schießt. Mit den Worten von J.L. Mackies bekannter Formulierung könnte man die wirkende Ursache (d.h. den Soldaten, der die Kanone abfeuert) als „insufficient but necessary part of a condition which is itself unnecessary but sufficient for the result“ bezeichnen.<sup>46</sup> Das heißt aber, dass Ursachen durch das in der hier vertretenen Verteidigung von Kants Kompatibilismus in Anspruch genommene kontrafaktische Abhängigkeitsverhältnis sehr viel zutreffender charakterisiert sind als durch die im Einwand zu Grunde gelegte Beschreibung als hinreichende Bedingungen.<sup>47</sup>

Kants allgemeine Auffassung von wirkenden Ursachen lässt sich wie folgt auf transzendental freie Handlungen übertragen: Wenn Kant die transzendental freie Aneignung  $H$  eines intelligiblen Charakters  $C_I$  als wirkende Ursache (*causa efficiens*) bezeichnet, die dazu führt, dass der Lügner einen bestimmten empirischen Charakter  $C_E$  hat und die Lüge  $L$  stattfindet, dann bedeutet das keineswegs, dass  $H$  allein dafür hinreichend ist, dass die Welt eine ist, in der  $L$  und  $C_E$  existieren, und dass der Handelnde allein dafür sorgt, dass die Welt eine ist, in der die Naturgesetze und der Weltverlauf dazu führen, dass  $L$  und  $C_E$  existieren.  $H$  ist genauso nur eine von vielen notwendigen und nur zusammengenommen hinreichenden Bedingungen für die Existenz von  $L$  und  $C_E$ , wie es das Anzünden der Kanonenkugel für deren Schießen ist. Genau in diesem Sinne beschreibt Kant die Rolle des freien Handelns der Vernunft in der Reflexion 5611: „Die Vernunft ist [...] in ansehung ihrer eignen Caussalitaet frey [...]. Wäre alles durch Vernunft bestimmt, so wäre alles nothwendig, aber auch gut. Wäre es durch die Sinnlichkeit bestimmt, so wäre nichts Böses oder Gutes, überhaupt nichts praktisches. Nun sind die Handlungen durch sinnlichkeit großen Theils veranlaßt, aber nicht gänzlich bestimmt; denn die Vernunft muß ein complement der Zulänglichkeit geben“ (AA XVIII 252; Hervorhebung von mir, T.R.).<sup>48</sup>

Auch in Kants veröffentlichten Schriften finden sich umfangreiche Hinweise darauf, dass er keineswegs der Meinung gewesen sein kann, dass die transzendental freie Handlung  $H$  eine hinreichende Bedingung für die Lüge  $L$ , den empirischen Charakter  $C_E$  oder den intelligiblen Charakter  $C_I$  ist:

<sup>46</sup> J. L. Mackie: *Causes and Conditions*. – In: *American Philosophical Quarterly*. Vol. 2/4 (1965). 245-264. 245.

<sup>47</sup> Ich hatte oben in der Diskussion des ‚altered past‘-Kompatibilismus gesagt, dass Kant als Vertreter dieser Position nicht annehmen dürfte, dass jede Art von kontrafaktischer Abhängigkeit hinreichend für Verursachung ist, weil er sonst die Möglichkeit eines kausalen Einflusses auf die Vergangenheit einräumen müsste. Auf der Grundlage der eben erläuterten Kantischen Ausführungen könnte man sagen, dass die Wahrheit eines kontrafaktischen Konditionals der Form „Wäre  $p$  nicht der Fall gewesen, dann wäre  $q$  nicht der Fall gewesen“ nur dann hinreichend dafür ist, dass der Sachverhalt, dass  $p$ , eine Ursache für den Sachverhalt, dass  $q$  ist, wenn er zusammen mit anderen Sachverhalten eine hinreichende Bedingung dafür ist, dass  $q$ . (So wie in Kants Beispiel der Soldat zusammen mit den verschiedenen *conditiones sine qua non* hinreichend für den Kanonenschuss ist.) Diese Bedingung ist im Falle eines kontrafaktischen Konditional wie „Wäre  $L$  nicht aufgetreten, wäre das-und-das Ereignis in der kausalen Vorgeschichte von  $L$  nicht aufgetreten“ nicht erfüllt.

<sup>48</sup> Die Formulierung „complement der Zulänglichkeit“ entspricht dem lateinischen „complementum ad sufficientiam“, das Kant in der oben zitierten Passage aus der *Religionslehre Politic* (AA XXVIII 1105) verwendet hatte, um zu beschreiben, was zu einer Ursache hinzukommen muss, damit das Ergebnis eine hinreichende Bedingung für die Wirkung ist.

(i) Im Fall von  $L$  ist dies besonders leicht zu zeigen, schließlich soll sich unsere Verantwortlichkeit für  $L$  dadurch ergeben, dass wir für unseren empirischen Charakter  $C_E$  verantwortlich sind,  $C_E$  allein kann aber natürlich nicht dazu führen, dass irgendeine Handlung stattfindet, sondern es bedarf bestimmter Handlungsumstände, damit die Substanz, die  $C_E$  hat, ihrem empirischen Charakter gemäß handelt.

(ii) Auch in dem Fall, in dem sich der Mensch in einer zeitlosen Handlung  $H$  einen bestimmten intelligiblen Charakter  $C_I$  verschafft und in seiner obersten Maxime die Triebfeder der Sittlichkeit der der Selbstliebe unterordnet, ist  $H$  nicht allein hinreichend dafür, dass er  $C_I$  hat. Die oben genannte „*conditio sine qua non*“, die als weitere Ursache vorhanden sein muss, besteht in diesem Fall in der Tatsache, dass der Mensch das moralische Gesetz und seine eigenen Neigungen als mögliche Bestimmungsgründe seines Handelns in sich vorfindet, ansonsten könnte er schließlich nicht das erstere den letzteren unterzuordnen. Dass diese Bestimmungsgründe in ihm vorhanden sind, hat sich der Mensch nicht frei ausgesucht: Das moralische Gesetz „dringt sich ihm vielmehr kraft seiner moralischen Anlage unwiderstehlich auf“ und er „hängt [...] auch vermöge seiner gleichfalls schuldlosen Naturanlage an den Triebfedern der Sinnlichkeit“ (AA VI 36). Diese beiden Faktoren sind im oben genannten Sinne „*concausae*“ der intelligiblen Handlung  $H$ , d.h. „mitwirkende Ursachen“, die weder alleine noch zusammengenommen hinreichend dafür sind, dass der Mensch einen bestimmten intelligiblen Charakter hat. Dies sind sie erst zusammen mit der Handlung  $H$ , weil erst diese festlegt, in welchem Bedingungsverhältnis sie zueinander stehen.

(iii) Schließlich kann man auch belegen, dass Kant nicht der Meinung gewesen sein kann, dass der intelligible Charakter eines Menschen eine hinreichende Bedingung für seinen empirischen Charakter ist. In der *Religionsschrift* sagt Kant, dass ein guter empirischer Charakter, d.h. einer, der den Menschen zu einem gesetzeskonform handelnden Wesen macht, Resultat eines bösen intelligiblen Charakters sein kann (AA VI 37), und er kritisiert, dass Menschen mit einem untadeligen Lebenswandel vorschnell annehmen, sie hätten einen guten intelligiblen Charakter, „ohne doch nachzuforschen, ob es nicht blos etwa Verdienst des Glücks sei, und ob nach der Denkungsart, die sie in ihrem Innern wohl aufdecken könnten, wenn sie nur wollten, nicht gleiche Laster von ihnen verübt worden wären, wenn nicht Unvermögen, Temperament, Erziehung, Umstände der Zeit und des Orts, die in Versuchung führen, (lauter Dinge, die uns nicht zugerechnet werden können) davon entfernt gehalten hätten“ (AA VI 38). Aus dieser Bemerkung folgt, dass der intelligible Charakter allein nicht hinreichend dafür sein kann, ob jemand einen empirischen Charakter hat, der zu gesetzeswidrigen Taten führt, schließlich kann selbst jemand mit bösem intelligiblen Charakter durch die richtige Erziehung oder durch glückliche Fügung böse empirische Handlungen vermeiden. Folglich besteht die Rolle des intelligiblen Charakters für die moralische Zurechenbarkeit unserer Handlungen nicht darin, eine hinreichende Bedingung

dafür zu sein, dass jemand einen bestimmten empirischen Charakter hat, aus dem sich zusammen mit den Handlungsumständen seine Handlungen ergeben. Die Zurechenbarkeit böser Handlungen ergibt sich vielmehr daraus, dass der Handelnde keinen bösen empirischen Charakter gehabt hätte, wenn er keinen bösen intelligiblen Charakter gehabt hätte, d.h. die Triebfeder der Sittlichkeit nicht der der Selbstliebe untergeordnet hätte.<sup>49</sup> Somit ist die Rolle des intelligiblen Charakters genau durch jenes kontrafaktische Abhängigkeitsverhältnis richtig beschrieben, das ich in meiner Verteidigung von Kants Kompatibilismus vorausgesetzt habe.<sup>50</sup>

---

<sup>49</sup> Kann Kant auf dieselbe Weise erklären, weshalb wir Menschen – anscheinend zu Recht – manchmal auch für ihre moralisch guten Handlungen verantwortlich machen und dafür loben? Unter Rückgriff auf das kontrafaktische Abhängigkeitsverhältnis zwischen der vermeintlich guten Handlung und der Aneignung einer bösen obersten Handlungsmaxime kann er dies jedenfalls nicht tun, denn bei einer guten Handlung ist es ja gerade nicht der Fall, dass sie nicht stattgefunden hätte, wenn diese Aneignung ausgeblieben wäre und der Mensch statt dessen die Triebfeder der Selbstliebe der der Sittlichkeit untergeordnet hätte. Man könnte Kant an dieser Stelle allerdings mit der bereits oben verteidigten Rekonstruktion seiner Lehre von der Unterordnung der Triebfedern in der obersten Handlungsmaxime zu Hilfe kommen (vgl. Anm. 16). Dieser Interpretation zufolge besteht die genannte Unterordnung im Falle eines bösen intelligiblen Charakters darin, dass es für die Menschen, die diesen Charakter haben, einen Grad der Verletzung ihres Eigeninteresses gibt, bei dem sie nicht mehr dem moralischen Gesetz folgen würden. Dieser Sachverhalt allein legt nicht fest, bei *welchem* Grad der Verletzung des Eigeninteresses dies für einen Menschen jeweils der Fall ist, bestimmt also den empirischen Charakter nicht vollständig. Kants obige Zitate machen klar, dass der besagte Grad zumindest teilweise von Faktoren abhängig ist, auf die der Mensch keinen Einfluss hat (wie Temperament und Erziehung). Wenn man nun annimmt, dass er aber *auch* vom intelligiblen Charakter des Menschen und seiner zeitlosen Aneignungstat abhängt und sich aus einem Zusammenspiel von diesen mit den nicht von ihm zu verantwortenden Faktoren ergibt, kann der Mensch auch für gute Handlungen verantwortlich gemacht werden. Solche Handlungen hätten nämlich nicht stattgefunden, wenn der Mensch sich durch seine intelligible Tat einen intelligiblen Charakter angeeignet hätte, bei dem die Latte für eine Abweichung vom Sitzensgesetz noch niedriger gelegen hätte.

<sup>50</sup> Ich habe frühere Versionen der hier entwickelten Interpretation in Vorträgen auf Tagungen in Göttingen, Saarbrücken und Miami vorgestellt. Für sehr hilfreiche Anregungen und Kritik möchte ich mich herzlich bei den Zuhörern dieser Vorträge sowie bei Wolfgang Ertl, Vera Flocke, Stefanie Grüne, Andreas Kemmerling, Franz Knappik und Thomas Krödel bedanken.

## Literatur

*Alexander Gottlieb Baumgarten: Metaphysica.* Übersetzt, eingeleitet und herausgegeben von Günter Gawlick und Lothar Kreimendahl. Stuttgart 2011.

*Jonathan Bennett: Kant's Theory of Freedom.* – In: *Allen Wood (Ed.): Self and Nature in Kant's Philosophy.* Ithaca 1984. 102–112.

*Jonathan Bennett: Counterfactuals and Temporal Direction* – In: *The Philosophical Review.* 93 (1984). 57-91.

*Jochen Bojanowski: Kants Theorie der Freiheit. Rekonstruktion und Rehabilitierung.* Berlin 2006.

*Wolfgang Ertl: 'Ludewig' Molina and Kant's Libertarian Compatibilism.* – Erscheint in: *Matthias Kaufmann et al. (Ed.): A Companion to Luis de Molina.* Kölln, Leiden. New York 2012.

*Wolfgang Ertl: Schöpfung und Freiheit. Ein kosmologischer Schlüssel zu Kants Kompatibilismus.* – In: *Norbert Fischer (Hg.): Kants Metaphysik und Religionsphilosophie.* Hamburg 2004. 43–76.

*Wolfgang Ertl: Hud Hudson: Kant's Compatibilism.* – In: *Kant-Studien.* 90 (1999). 371–384.

*Hud Hudson: Kant's Compatibilism.* Ithaca, London 1994.

*David Lewis: Counterfactual Dependence and Time's Arrow.* – In: *Nous.* 13 (1979). 455–76.

*David Lewis: Are We Free to Break the Laws?.* – In: *Theoria.* 47 (1981). 113–121.

*J. L. Mackie: Causes and Conditions.* – In: *American Philosophical Quarterly.* Vol. 2/4 (1965). 245-264.

*Derk Peereboom: Kant on Transcendental Freedom.* – In: *Philosophy and Phenomenological Research* 73 (2006). 537–567.

*Tobias Rosefeldt: Dinge an sich und sekundäre Qualitäten.* – In: *Jürgen Stolzenberg (Hg.): Kant in der Gegenwart.* Berlin, New York 2007. 167–209.

*Ralph Walker: Kant.* London, Henley, Boston 1978.

*Erik Watkins: Kant and the Metaphysics of Causality.* Cambridge 2005.

*Marcus Willaschek: Praktische Vernunft. Handlungstheorie und Moralbegründung bei Kant.* Stuttgart, Weimar 1992.

*Michael Wolff: Freiheit und Determinismus bei Kant.* – Erscheint in: *B. Tuschling & W. Euler (Hg.), Kants Metaphysik der Sitten. Philosophische und editorische Probleme.* Berlin (im Erscheinen).

*Allen Wood: Kant's Compatibilism.* – In: *Allen Wood (Ed.): Self and Nature in Kant's Philosophy.* Ithaca 1984. 73–101.

*Simon Shengjian Xie: What Is Kant: A Compatibilist or an Incompatibilist? A New Interpretation of Kant's Solution to the Free Will Problem.* – In: *Kant-Studien* 100 (2009). 53–76.