

oder als indeterminiert und damit unkontrollierbar zu haben, von dem infiniten Regress ganz abgesehen, den der Rekurs auf den Willen hier offensichtlich eröffnen würde.

Der Sonderweg Kants, die Rücknahme des Freiheitsgrundes aus der Welt in die „intelligible“ Sphäre reiner Begriffe, scheitert nicht nur an seinen inneren Inkonsistenzen, sondern vor allem an der Unmöglichkeit, von einer solchen transzendentalen Freiheit aus eine kausale Brücke in die Welt der „Erscheinungen“ zu schlagen. Wer will mag einen Begriff „noumenaler Freiheit“ aus dem sittengesetzlichen „Factum der Vernunft“ deduzieren. Aber „Kausalität aus Freiheit“ *dieser* Provenienz ist eine Fata Morgana.

Das Resultat unserer Analysen zum libertären Inkompatibilismus lässt sich nicht bündiger fassen als in die Formel eines seiner bedeutendsten Vertreter: „Free will remains a mystery“.<sup>125</sup> Ein genauerer Blick zeigt den prinzipiellen Grund: eine Art Hase-und-Igel-Verhältnis zwischen dem Willen und seiner neuronalen Basis. Welche gedankliche Position mittels welcher begrifflichen Finessen man für den mentalen Hasen des Wollens oder Entscheidens auch reklamieren möchte, um dessen Freiheit zu behaupten – der Igel des Neurophysiologischen ist immer schon da. Denn ohne ihn, ohne es, ist überhaupt nichts Mentales vorstellbar, ein Wille sowenig wie irgend ein sonstiger Vorgang des Bewusstseins.<sup>126</sup> Daher scheint es unumgänglich, sich bei der Begründung von Freiheit genau damit zu arrangieren.

## V. Die kompatibilistischen Positionen: Grundlagen und Grenzen

Das tun die Kompatibilisten. Sie erkennen sowohl die Unlösbarkeit des Zusammenhangs von Gehirn und „Geist“ als auch einen dabei

125 *van Inwagen* (Anm. 116); gegen ihn wenig überzeugend *Griffith*, *Does Free Will Remain a Mystery?*, in: *Philosophical Studies* 124 (2005), 261 ff.

126 Genau besehen nicht einmal sein akausales Dasein als Noumenon in einer Welt reiner Vernunft: Auch diese ist nicht ohne Denken und dieses seinerseits nicht ohne seinen „unreinen“ Ursprung im Gehirn vorstellbar.

(irgendwie) maßgeblichen Vorrang des Gehirns an. Der erste gemeinsame Nenner ihrer Theorien ist deshalb die Zurückweisung der Annahme, der Wille sei in irgendeinem Sinn souverän gegenüber seinen neuronalen Grundlagen. Vielmehr deuteten alle verfügbaren empirischen Evidenzen auf einen determinierten Zusammenhang zwischen beiden.<sup>127</sup> Doch bedrohe dies, und das ist der zweite gemeinsame Nenner, weder eine richtig verstandene Freiheit noch die Möglichkeit der Zuschreibung von Verantwortung und Schuld. Will man das überprüfen, so muss man sich zunächst über diesen Zusammenhang von Geist und Gehirn, von Mentalem und Neuronalem Klarheit verschaffen.

### *1. Das Gehirn-Geist-Problem und seine Bedeutung für die Frage der Willensfreiheit*

Wenn „der Geist“, nämlich alle mentalen Phänomene, irgendwie vom Funktionieren des Gehirns abhängig ist, was könnte dann noch die Bezeichnung „freier Wille“ verdienen? Unsere bisherigen Erörterungen haben schon erkennen lassen, dass diese Frage zwei Seiten hat:

(1) *Das Problem der autonomen Urheberschaft:* Kann ein derart vom physikalischen System des Gehirns hervorgebrachter Wille noch als hinreichend autonome Äußerung einer freien Person und damit als Grundlage ihrer Verantwortlichkeit begriffen werden?

(2) *Das Problem der mentalen Verursachung:* Wie kann ein geistiges Phänomen wie der Wille (oder eine Entscheidung) Wirkungen auf die physische Welt des Körpers ausüben, insbesondere Handlungen auslösen und steuern?

#### *1.1 Identitätstheorien*

Viele Philosophen glauben, beide Fragen ließen sich auflösen, wenn man eine bestimmte Form der Identität von Geist und Gehirn anneh-

127 Zur Erinnerung: „determiniert“ meint hier keinen strikt gesetzlichen Determinismus, sondern nur einen Zusammenhang stabiler, naturgegebener Regularitäten, die für Menschen nicht veränderbar sind; vgl. oben II.

me.<sup>128</sup> „Identität“ heißt hier nicht „Ununterscheidbarkeit“ im Sinne der Logik; gemeint ist vielmehr die *Reduzierbarkeit* des Mentalen auf das Neurophysische. Einigen Autoren geht es dabei lediglich um eine epistemische These, nämlich um die logische Reduktion mentalistischer Beschreibungen auf neurophysiologische. Andere vertreten einen weiterreichenden, ontologischen Reduktionismus: Die mentalen Eigenschaften selber seien ohne Rest auf Eigenschaften des Gehirns zurückzuführen, nämlich aus diesen empirisch vollständig erklärbar.<sup>129</sup> Ein einfaches Modell mag das Gemeinte verdeutlichen: Man stelle sich ein Raster mit einer Million winziger Punkte (sog. Pixels) vor, deren jeder entweder hell oder dunkel ist. In ihrer Gesamtheit ergeben sie das naturgetreue Panorama einer Stadt mit einer Fülle spezifischer Details. Dieses Stadtbild und alle seine besonderen Eigenschaften (Formen, Konturen, Linien) existieren selbstverständlich wirklich. Will man aber wissen, wie und warum sie existieren, so wird eine zutreffende Erklärung ihre Existenz vollständig auf das spezifische Arrangement der hellen und dunklen Pixels reduzieren.<sup>130</sup> Reduktionistische Theorien verfolgen also nicht, wie manchmal befürchtet wird, das absurde Anliegen, irgendwelche Dinge der Welt zu „eliminieren“. Was sie eliminieren, ist lediglich die Annahme einer ontologischen Unabhängigkeit dieser Dinge in einer wahren Erklärung der Welt. Genau dieses Anliegen verfolgen Reduktionisten auch für die Erklärung des „Geistes“.

128 Das ist ersichtlich eine starke physikalistische Position (s. oben unter IV.1.2, S. 32 f.). Man muss aber, wie wir noch sehen werden, als Physikalist kein Identitätstheoretiker sein; es gibt auch einen nicht-reduktiven Physikalismus.

129 Aus der unüberschaubaren Flut der Literatur zum Reduktionismus nur *E. Nagel*, *The Structure of Science*, 1961, S. 336 ff. (zur allgemeinen Wissenschaftstheorie); zum psychoneuralen Reduktionismus nur *Crooks*, *Intertheoretic Identification and Mind-Brain Reductionism*, in: *The Journal of Mind and Behavior* 23 (2002), 193 ff. (*contra* Redukt.; mit vier kritischen Kommentaren), sowie *D. Lewis*, *Reduction of Mind*, in: *Guttenplan* (ed.), *A Companion to the Philosophy of Mind*, 1994, S. 412 ff. (*pro* Redukt.); umfassend *Bickle*, *Psychoneural Reduction: The New Wave*, 1998. – Ursprung der Debatte zur psychoneuralen Identität war *Smart*, *Sensations and Brain Processes*, in: *Philosophical Review* 69 (1959), 141 ff. (*pro* Identität).

130 Beispiel von *D. Lewis* (Anm. 129), S. 413.

Es ist leicht zu sehen, dass für Identitätstheoretiker die Frage (1) nach einer autonomen Urheberschaft an der Willensentstehung weitgehend obsolet ist: Unabhängig vom Gehirn sei so etwas nicht vorstellbar, weil der Wille eben mit bestimmten neuronalen Prozessen identisch sei. Doch berühre das eine vernünftig verstandene Autonomie der Person nicht. Schließlich sei es *ihr* Gehirn, dessen Funktionieren ihren Willen hervorbringe. Dass es dabei naturgegebenen Regularien folge, die keiner Willenssteuerung, vielmehr nur der Selbststeuerung des Gehirns unterlägen, mache den Willen nicht unfrei, wenn nur das Gehirn selber ohne manifesten neuro-psychologischen Defekt funktioniere. Im Gegenteil: diese Identität des Willens mit seinen neuronalen „Realisierern“ verbürge gerade das Authentische, Höchstpersönliche seiner Beschaffenheit. Denn alles, was den Willenden selbst ausmache, seine Charakterzüge, Präferenzen, kognitiven, emotionalen und motivationalen Eigenschaften, residiere in der neuronalen Architektur seines Gehirns. In Entscheidungssituationen werde es durch sensorische „Inputs“ umfassend aktiviert, integriert, synchronisiert, und führe genau deshalb zur Bildung eines höchstpersönlichen, zurechenbaren Willens.

Dieses Argument von der Authentizität des Willens, die gerade durch dessen neuronale Determination verbürgt werde, steht ersichtlich nicht nur Identitätstheoretikern offen. Plausibel beanspruchen kann es auch, wer lediglich einen stabilen, neuronal erzeugten und naturgegebenen Regularien folgenden Zusammenhang zwischen Geist und Gehirn annimmt, ohne doch deren Identität zu behaupten. Auf der Hand liegt freilich, dass allein damit die Frage nach Verantwortlichkeit und Schuld nicht zu beantworten ist. Ich komme darauf zurück.

Doch selbst, wenn man sich mit einem so explizierten Freiheitsbegriff zufrieden gibt, ist damit von den beiden oben genannten Grundproblemen allenfalls das erste, das der autonomen Urheberschaft, gelöst. Identitätstheoretiker behaupten nun, ihre Grundthese bringe auch das zweite, das der mentalen Verursachung, zum Verschwinden. Denn wenn das Mentale mit seiner neuronalen Grundlage identisch sei, dann sei die vermeintlich mentale Verursachung, deren Möglichkeit so unbegreiflich erscheine, in Wahrheit eine neuronale Verursachung.

Und an *deren* Möglichkeit als der eines Kausalzusammenhangs zwischen zwei physischen Ereignissen zweifelt niemand; die spezifische Ursächlichkeit neuronaler Erregungsmuster für Körperbewegungen stehe empirisch ohnehin außer Zweifel. Darüber hinaus sprächen für die Identitätstheorie, so behaupten ihre Anhänger, weitere entscheidende Gründe, die von ihrem Potential zur Lösung der beiden genannten Grundfragen unabhängig seien. Insbesondere passe sie von allen verfügbaren Konzeptionen am besten zu unserem wissenschaftlichen Weltbild. So wie die mittlere Molekularbewegung in einer Flüssigkeit deren Temperatur einfach *sei* und diese nicht erst auf geheimnisvollem Wege erzeuge, oder wie ein Blitz eine elektrische Entladung sei und nichts darüber hinaus, und wie ganz allgemein makrophysikalische Eigenschaften eines Gegenstands in dessen atomarer Struktur unmittelbar realisiert seien und nicht erst (zusätzlich) von ihr erzeugt würden – ganz genauso seien mentale Ereignisse unmittelbar identisch mit den sie realisierenden neuronalen Ereignissen.

### 1.1.1 *Mentale Verursachung?*

Stimmt das? Für und gegen die Identitätstheorie sind in den vergangenen Jahrzehnten zahllose Argumente in Stellung gebracht worden, die hier auf sich beruhen müssen.<sup>131</sup> Zwei aus der Gruppe der Gegenargumente erscheinen mir aber durchschlagend. Das erste bestreitet die oben skizzierte Lösung zum Problem der mentalen Verursachung; das zweite verwirft aus einem prinzipiellen Grund die Identitätsannahme im ganzen. Zum letzteren, prinzipiellen Argument komme ich im nächsten Abschnitt (1.1.2). Der Einwand gegen die behauptete Lösung des Kausalitätsarguments lautet so: Ereignisse verursachen andere Ereignisse nicht einfach schlechthin und *tout court*, sondern jeweils vermöge bestimmter Eigenschaften. Diesen komme daher im Rahmen des Verursachungsvorgangs die eigentliche kausale Relevanz zu,

131 Gute Übersicht bei *Kim*, *Philosophy of Mind*, 1996, S. 47 ff. („Mind as Brain“); ausgezeichnet *Walter*, *Mentale Verursachung. Eine Einführung*, 2006; instruktiv auch *Heil/Mele* (eds.), *Mental Causation*, 1993.

während andere Eigenschaften des ursächlichen Ereignisses für dessen Kausalwirkung bedeutungslos sein könnten. Nehmen wir an, die Sopranistin S singt vor einem Fenster im hohen C und in einer bestimmten Lautstärke das Wort „Zerspring!“, woraufhin die Scheibe kaputtgeht.<sup>132</sup> Es ist völlig korrekt zu sagen, das Ereignis des Singens dieses Wortes habe die Scheibe zerspringen lassen. Aber für eine gründliche Beschreibung ist es zu ungenau. Dafür müsste es lauten: Nicht vermöge der Bedeutung des gesungenen Wortes, sondern vermöge der dabei angeschlagenen Tonhöhe und Lautstärke ist das Singen für die Zerstörung der Scheibe kausal geworden; hätte die Sängerin in gleicher Höhe und Lautstärke „Bleib heil!“ gesungen, so wäre die Scheibe genauso kaputtgegangen.

Im gleichen Sinne kausal irrelevant, so lautet nun der Einwand gegen die Identitätstheorie, ist aber (zumindest möglicherweise) das mentale Ereignis des Willens oder Entschlusses zu einer Handlung, auch wenn es mit den neuronalen Vorgängen, durch die es realisiert wird, identisch sein sollte – so wie in unserem Beispiel das Ereignis des Singens des Wortes „Zerspring!“ identisch ist mit dem Ereignis des Singens eines Tons in bestimmter Höhe und Lautstärke, und dennoch nichts Kausales zum Zerspringen der Scheibe beiträgt. Kurz: nicht vermöge seiner mentalen, sondern nur vermöge seiner physischen Eigenschaften könne das psycho-physische Gesamtereignis „Willensentschluss“ kausale Wirkungen in der physischen Welt hervorrufen, z.B. die nachfolgende Handlung einleiten. Daher löse die Identitätstheorie entgegen ihrem Anspruch das Problem der mentalen Verursachung nicht.

Ich halte den Einwand für richtig. Die Frage, wie es möglich ist, dass etwas Mentales, also *per definitionem* Nichtkörperliches, wie der Wille kausale Wirkungen in der Körperwelt hat, erledigt sich nicht dadurch, dass man es einfach mit einem physischen (kausalitätsfähigen) Ereignis identifiziert. Freilich bleibt dann das Problem der mentalen Verursachung in seiner ganzen irritierenden Rätselhaftigkeit

132 Vielzitiertes Veranschaulichungsbeispiel von *Dretske*, *Explaining Behaviour: Reasons in a World of Causes*, 1988, S. 79; sowie *ders.*, *Reasons and Causes*, in: *Philosophical Perspectives* 3 (1989), 1 f.

insgesamt ungelöst zurück. Kapitulierte man vor ihm und erkennt nur physische Verursacher bzw. nur ein Verursachen vermöge physischer Eigenschaften an, dann scheint das gesamte mentale Erleben des Menschen, also das, was jedes individuelle Leben für seinen Inhaber erst lebenswert macht, zur bloßen Begleitmusik des Physischen zu werden: zum Epiphänomen der neuronalen Vorgänge in seinem Gehirn, von diesen erzeugt, aber ohne jede kausale Rückwirkung auf sie (ja auf *irgendetwas* in der Welt).<sup>133</sup> Das ist nicht nur in hohem Maße kontraintuitiv, sondern außerdem ein Gedanke, den viele Philosophen schwer erträglich finden. Wenn Mentales niemals für physische Vorgänge kausal werden kann, dann kann auch kein physischer Vorgang mit etwas Mentalem erklärt werden – also z.B. keine Handlung mit dem Willen des Handelnden. Das erscheint vielen absurd, und müssten sie es dennoch zugestehen, erschiene es ihnen katastrophal. Der amerikanische Philosoph Jerry Fodor hat der entsprechenden Antipathie plastischen Ausdruck gegeben:

„Wenn es nicht buchstäblich wahr ist, dass ich nach etwas greife, *weil* ich es haben möchte, dass ich mich kratze, *weil* es mich juckt, und dass ich etwas sage, *weil* ich es glaube..., wenn nichts davon buchstäblich wahr ist, dann ist praktisch alles falsch, was ich über irgendetwas glaube, und das ist das Ende der Welt.“<sup>134</sup>

Vielleicht ein wenig übertrieben, denn unser subjektives Erleben müsste sich ja allein durch die Einsicht in seinen epiphänomenalen Charakter nicht unbedingt selber verändern. Aber die Mischung aus unwirschem und beschwörendem Tonfall in Fodors Satz steht exemplarisch für viele Abhandlungen, die in den vergangenen Jahrzehnten

- 133 Einer der ersten Exponenten des Epiphänomenalismus war *Thomas Huxley*, der die kausale Wirkungskraft des Mentalen mit dem Pfeifen einer Dampflokomotive verglich, das von deren Betrieb erzeugt werde, aber keine Rückwirkung auf ihn habe; s. *Huxley*, *On the hypothesis that animals are automata, and its history* (1874), in: *Collected Essays*: Vol I, 1893, S. 199 ff., (240); Huxley begründete seine These für Tiere, hielt sie aber auch für den Menschen für gültig (S. 244); gegen Huxley *W. James*, *Are we automata?*, in: *Mind* 4 (1879), 1 ff.
- 134 *Fodor*, *Making Mind Matter More*, in: *ders.*, *A Theory of Content and Other Essays*, 1990, S. 137 ff. (156) (Hervorh. von mir); s. auch *McLaughlin*, *Epiphenomenalism*, in: *Guttenplan* (Anm. 129), S. 277 ff.

zum Problem des Epiphenomenalismus erschienen sind: Zumeist wurde dessen Abwegigkeit fraglos vorausgesetzt und nur ein entsprechender Nachweis gesucht. Gleichwohl blieb das Problem gegenwärtig und renitent. Denn wie eine mentale Verursachung denkbar sei, das war und ist trotz der Aura des Absurden, die den Epiphenomenalismus umgibt, so rätselhaft wie eh und je. Eine rundum überzeugende Lösung des Problems ist auch heute, nach unzähligen Erklärungsversuchen, nicht in Sicht. Anfang 2006 erschien im renommierten „Journal of Consciousness Studies“ ein spezielles Themenheft unter dem Titel „Epiphenomenalism“, dem die Herausgeber in ihrem Vorwort die sprechende Ergänzung „Dead End or Way Out?“ beifügten. Das Meinungsspektrum der acht Beiträge spiegelt genau diese Frage als aktuelles und ungelöstes Problem der Philosophie des Geistes wider.<sup>135</sup> Und damit zugleich das fortbestehende Rätsel seines Zwillingproblems, der Frage nach einer Möglichkeit mentaler Verursachung. Vor diesem Hintergrund darf man es erstaunlich finden, dass im Strafrecht das Problem der mentalen Verursachung bislang so gut wie vollständig ignoriert worden ist.<sup>136</sup> Ein Freiheitsproblem sieht man allenfalls in der Entstehung, nicht aber in der möglichen Wirkung des Willens. Danach gefragt wird deshalb regelmäßig erst im Rahmen der Schuldprüfung, nicht dagegen bereits beim Handlungsmerkmal im Tatbestand. Aber schon dort gibt es eines: das des „willensgesteuerten“ Verhaltens. Dessen Vorliegen, Grundbedingung einer Zuständigkeit des Strafrechts überhaupt, wird für den Normalfall bewussten menschlichen Handelns regelmäßig ohne weiteres bejaht. Das könnte sich als voreilig erweisen. Gibt es so etwas überhaupt? Was hieße es, wenn die Frage verneint werden müsste? Wäre ein *nicht* handlungssteuernder Wille oder eine nicht willensgesteuerte Handlung „frei“?

135 Pauen/Staudacher/Walter (eds.), Epiphenomenalism, Journal of Consciousness Studies 13, No 1-2, 2006; zur Offenheit des Rätsels (nach 250 Seiten gründlicher Analyse) auch Walter (Anm. 131), S. 258 f.

136 Eine Ausnahme ist Detlefsen (Anm. 10), S. 309 ff., 324 f., die (allerdings ohne Blick auf die philosophische Debatte) das Problem der mentalen Verursachung deutlich sieht, deren Möglichkeit – neben einer vorausgesetzten neuronalen Determination – skeptisch beurteilt, gleichwohl aber den Epiphenomenalismus ablehnt.

Und bejaht man Handlungssteuerung durch den Willen dennoch: wie genau macht er das? Wie kann er als etwas Immaterielles materielle Wirkungen erzeugen ohne den Energieerhaltungssatz zu verletzen? Und wenn er's könnte: warum braucht er gleichwohl stets ein funktionierendes Gehirn, um eine körperliche Handlung auszulösen? Wie „steuert“ er diese dann, wenn doch das Funktionieren des unmittelbar auslösenden Gehirns naturgesetzlichen Regularien folgt? Und warum kann mein Geist, wenn er wirklich *so etwas* kann, immer nur auf mein Gehirn und auf kein anderes einwirken?

Auf keine dieser Fragen gibt es eine überzeugende Antwort *pro libertate*. Nicht nur gibt es keine Spur eines empirischen Indizes für unmittelbare Wirkungen des Geistes auf das Gehirn, es gibt im Gegenteil gute Gründe für die Annahme, dass solche Wirkungen unmöglich sind. Das schließt es ja keineswegs aus, das Strafrecht dennoch für tatbestandliches Handeln zuständig zu machen. Man wird aber fragen dürfen, ob die Strafrechtsdogmatik ihre schlafwandlerische Selbstsicherheit im Umgang mit dem Grundbegriff der „willensgesteuerten Handlung“ nicht doch vorübergehend gegen eine aufgewecktere Skepsis eintauschen sollte. Nicht jeder dogmatische Begriff, der in seiner Buchstäblichkeit keinen fassbaren Inhalt (mehr) hat, muss ja sofort abgeschafft werden, und der der „Willenssteuerung“ wäre bei weitem nicht der erste und einzige dieser Art im geläufigen Inventar des Strafrechts. Was aber nötig wäre, ist eine größere Klarheit darüber, was daran sachlich glaubiger Sinn und was bloß noch überlieferte Metaphorik ist.

### 1.1.2 *Qualia, oder: Was Mary nicht wissen konnte*

Der zweite Einwand gegen die Identitätstheorie ist radikaler: Ein reduktiver (identitärer) Physikalismus, der alle Erscheinungen der Welt für (im Prinzip) erklärbar hält durch ihre analytische Rückführung auf ein letztes, regelmäßige molekulares oder atomares Fundament des Physischen, sei falsch. Wohl könne man mittlerweile die meisten der traditionell rätselhaften Phänomene, z.B. die Entstehung von Leben aus „toten“ materiellen Bausteinen, physikalistisch-reduktiv erklären, nicht anders als den Druck eines Gases durch dessen mittlere Mole-

külgeschwindigkeit.<sup>137</sup> Aber eines dieser Probleme bleibe für jeden denkbaren wissenschaftlichen Physikalismus unerreichbar: die Frage nach dem Zusammenhang von Geist und Gehirn. Das lasse sich zeigen: „Ein Baum im Wald stürzt um. Ist jemand in der Nähe und hört dies, dann gibt es das Geräusch eines stürzenden Baumes; andernfalls gibt es nur Schallwellen. Wenn nun jemand den Baum stürzen hört, wo genau steckt dann dieser Unterschied – das subjektive Erfahren des Geräusches? Öffne den Kopf des Hörenden, untersuche sein Gehirn in jeder einzelnen Windung; du wirst weder das Geräusch noch die Erfahrung des Geräuschhörens finden.“<sup>138</sup>

Hier wird die Idee plastisch. Nicht auf Physi(kali)sches reduzierbar sei der spezifisch subjektive Charakter des inneren, mentalen Erlebens. Der starke (reduktive) Physikalismus sei daher falsch.<sup>139</sup> Eine Beschreibung des gesamten Universums müsste als ein irreduzibles Grundelement im Bau der Welt das subjektive Erleben des Mentalen enthalten. Anders gewendet: eine rein physikalische Beschreibung des Inventars der Welt wäre notwendig unvollständig.

Traditionell unterscheidet man zwei verschiedene Typen von mentalen Zuständen: intentionale und phänomenale.<sup>140</sup> Zum ersteren Typus gehören innere Zustände „über etwas“ oder „gerichtet auf etwas“, wie nachdenken, wünschen, hoffen, lieben, hassen, etc.; zu ihnen gehört auch, leicht erkennbar, der Wille oder Entschluss zu einem bestimm-

137 Dazu *McLaughlin*, *The Rise and Fall of British Emergentism*, in: *Beckermann/Flohr/Kim* (eds.), *Emergence or Reduction?*, 1992, S. 49 ff.; instruktiv auch *Papineau* (Anm. 41). – Die manchmal in diesem Kontext genannten Inhalte des „objektiven Geistes“ oder *Karl Poppers* „dritter Welt“ (s. *ders.*, *Objektive Erkenntnis*, 1973, S. 88 f., *passim*), also etwa Normen, Religionen, Kulturwerte etc., sind übrigens ohne weiteres reduktiv-physikalistisch erklärbar. Was sich mit einer solchen Erklärung anfangen lässt (und v.a. was nicht), ist eine andere Frage.

138 *Glymour*, *A Mind is a Terrible Thing to Waste – Critical Notice: Jaegwon Kim, Mind in a Physical World*, in: *Philosophy of Science* 66 (1999), 455 ff.

139 *Thomas Nagel*, der mit seinem berühmten Aufsatz „What Is It Like to Be a Bat?“ (*Philosophical Review* 83 [1974], 435 ff.) das physikalistisch Unerklärbare des subjektiven Erlebens zum Thema gemacht hat, hielt freilich am Physikalismus fest, weil er eine physikalistisch-reduktive Erklärung anhand künftiger wissenschaftlicher Einsichten nicht ausschließen wollte.

140 S. bereits oben, zu und in Anm. 54.

ten Handeln. Der zweite Typus umfasst u. a. alle Sinneseindrücke: Empfindungen von Hitze, Kälte, Schmerz, Eindrücke beim Sehen, Hören, Riechen, Schmecken, etc. Die Begründungsstrategien gegen den reduktiven Physikalismus stützen sich meist allein (oder hauptsächlich) auf das Irreduzible der phänomenalen, nicht aber der intentionalen Zustände. Berühmt geworden ist ein entsprechendes Gedankenexperiment des australischen Philosophen Frank Jackson aus dem Jahr 1982:

Mary, eine exzellente Neurowissenschaftlerin, ist von Geburt an gezwungen, subjektiv in einer schwarz-weißen Welt zu leben. Denn sie hat einen Augendefekt, der jedes farbliche Sehen ausschließt. Nach langen Studien verfügt sie – so sei unterstellt – über *alle* physikalischen und neurobiologischen Informationen über *sämtliche* Vorgänge, die in einem Menschenkopf beim Sehen der Farbe Rot stattfinden: von der Reaktion der Retina auf das eintreffende Licht bis zu den neuronal-molekularen Vorgängen und den „feuernden“ Synapsen bei der Verarbeitung im Gehirn. Alle diese Vorgänge kennt sie auch aus zighfacher Beobachtung in Experimenten. Irgendwann kann Marys Augendefekt medizinisch behoben werden. Nun *sieht* sie zum erstenmal selbst die Farbe Rot.<sup>141</sup>

Jacksons Pointe ist, dass Mary nun etwas prinzipiell Neues lernt, etwas, das ihr kein noch so vollständiges Wissen über alle neurophysiologischen Vorgänge des Rotsehens vorher habe vermitteln können. Dieses kategorial Neue ist das Subjektiv-Qualitative oder Phänomenale des Sinneseindrucks (technisch: „Quale“, im Plural „Qualia“). Kenntnis davon sei nur mit dem Haben des mentalen Zustands selber, aber mit keinem Wissen über dessen neuronale Grundlagen zu erlangen. Der reduktive bzw. identitäre Physikalismus sei daher falsch.

Jacksons Gedankenexperiment hat eine Flut von Aufsätzen, zustimmenden wie ablehnenden, zum Problem der Qualia und des Physikalismus ausgelöst.<sup>142</sup> Diese Diskussion, die profunde Einsichten in das

141 Jackson, Epiphenomenal Qualia, in: *Philosophical Quarterly* 32 (1982), 127 ff.; ich habe das Szenario des Gedankenexperiments etwas verändert; das Jacksonsche ist deutlich bizarrer als das hier gewählte (eine solche vollständige Farbenblindheit, den sog. Achromatismus, gibt es tatsächlich); die gedankliche Pointe ist identisch.

142 Gute Dokumentation der Debatte in Ludlow/Nagasawa/Stoljar, *There's Something*

Problem erbracht, es aber nicht annähernd konsensfähig gelöst hat, muss hier auf sich beruhen. Zumindest in ihrem Kern halte ich Jacksons Überlegung für richtig.<sup>143</sup> Und in eben diesem Ausmaß erscheint der reduktive Physikalismus daher falsch.

Was bedeutet das für unsere Frage nach der Freiheit des Willens? Jedenfalls dies: Wir haben zumindest einige gute Gründe, Willensakte nicht für identisch mit ihren neuronalen Grundlagen zu halten. Zwar demonstriert Marys Beispiel dies unmittelbar nur für phänomenale Zustände (Qualia), nicht aber für intentionale (wie den Entschluss zu einer Handlung). Doch ist schon zweifelhaft, ob Qualia und intentionale Zustände stets klar voneinander unterschieden werden können. Jedenfalls erscheint es nicht unvernünftig, aus dem Umstand, dass Qualia nicht auf ihre neurophysiologischen Grundlagen reduzierbar sind, den Schluss zu ziehen, dass *kein* subjektiv-mentaler Zustand dies ist.<sup>144</sup> Der reduktive Physikalismus gibt uns daher auf unsere beiden Grundfragen, die nach einem hinreichend autonomen Ursprung des Willens und die nach dessen kausaler Wirkung auf Einleitung und Kontrolle von Handlungen, keine überzeugende Antwort. Und ebenso

About Mary, 2004. Angegriffen hat man v.a. Jacksons „knowledge argument“, die Behauptung, Mary erwerbe im Moment ihres erstmaligen Rotsehens eine bislang nicht vorhandene *Kenntnis von etwas*: einer (neuen) Tatsache. Ein phänomenaler Sinneseindruck sei keine Tatsachenkenntnis; Mary lerne also nichts Neues über die Welt; sie werde nur mit einem veränderten Modus des Sehens vertraut; der reduktive Physikalismus sei daher nicht widerlegt. Mir erscheint der Einwand weder triftig noch richtig: Mary lernt, dass *andere* Menschen beim Sehen von Rot etwas Ähnliches empfinden wie sie selbst im Moment (gewiss eine Tatsache). Überzeugend für ein (modifiziertes) „knowledge argument“ *M. Nida-Rümelin, On Belief About Experiences. An Epistemological Distinction Applied to the Knowledge Argument...*, in: *Philosophy and Phenomenological Research* 58 (1998), 51 ff.

143 Jackson selbst tut dies übrigens nicht mehr; er ist zum reduktiven Physikalismus zurückgekehrt, s. *ders.*, Postscript on Qualia, in: *Ludlow et al. (Anm. 142)*, S. 417 ff., sowie *ders.*, *Mind and Illusion*, ebda., S. 421 ff.

144 Ist z.B. das innere Ringen um eine schwere Entscheidung nur als intentionaler Zustand fassbar, oder „fühlt es sich“ nicht doch auch „irgendwie an“? Dass man Qualia und intentionale Zustände trennen und jedenfalls die letzteren (z.B. den „Willen“) physikalistisch-reduktiv erklären könne, meint aber *Kim, Mind in a Physical World*, 1998, S. 118 ff.; *ders.*, *The Causal Efficacy of Consciousness*, in: *Velmans/Schneider (Anm. 62)*, S. 406 ff.

wenig überzeugend ist sein Vorschlag, diese Fragen hinter immer genaueren physikalistischen Beschreibungen der neuronalen Vorgänge im Gehirn eines Handelnden gänzlich verschwinden zu lassen.

## 1.2 Andere Lösungen: Moderate Dualismen – Emergenz? Supervenienz?

Wer den vollständigen Reduktionismus, die Identität von Geist und Gehirn, verneint, kann dennoch Physikalist sein und daher die These einer souveränen Unabhängigkeit des Willens vom Gehirn ganz genauso verwerfen. Das tut der nicht-reduktive Physikalismus. Auch er hat verschiedene Spielarten.<sup>145</sup> Ihre gemeinsame Grundthese, dass sich das Mentale vom Neuronalen irreduzibel unterscheidet, führt freilich (erstens) zwingend zu irgendeiner Form von Dualismus zwischen „Leib und Seele“ und (zweitens) zu der Notwendigkeit, den postulierten besonderen Zusammenhang zwischen Gehirn und Geist wenigstens begrifflich plausibel zu machen.

Der Dualismus-Vorwurf ist zwar beliebt, aber langweilig.<sup>146</sup> Dass jemand, der einen Identitätsmonismus ablehnt, irgendeine Form von Dualismus akzeptieren muss, ist trivial und nicht im mindesten zu beanstanden. Viele setzen einen Monismus der Substanz (Gehirn) und einen Dualismus der Eigenschaften dieser Substanz (neuronal / mentale) voraus. An Benennungen hängt nicht viel; ein solcher Eigenschaftsdualismus ist jedenfalls kein unplausibler Ausgangspunkt für weitere Klärungen.

145 Ausgezeichnete Sammelbände dazu sind *Beckermann/Flohr/Kim* (Anm. 137); *Alter/Walter* (eds.), *Phenomenal Concepts and Phenomenal Knowledge. New Essays on Consciousness and Physicalism*, 2007; *Freeman* (ed.), *Consciousness and its Place in Nature*, 2006 (zu der verblüffenden Theorie *G. Strawsons*, dass Physikalisten das Mentale nur dann erklären könnten, wenn sie zugleich einen Panpsychismus annähmen).

146 Er knüpft an den diskreditierten Substanzdualismus von „res cogitans“ und „res extensa“ bei *Descartes* an (s. Anm. 30); einen solchen muss (sollte) aber kein Dualist heute noch vertreten. Für einen modernen, wissenschaftlich aufgeklärten Dualismus *Chalmers*, *Naturalistic Dualism*, in: *Velmans/Schneider* (Anm. 62), S. 359 ff.

Damit erhebt sich die Frage, wie der *Zusammenhang* zwischen diesen Eigenschaften bzw. (konkret für unser Thema) zwischen dem Willen und seinen neuronalen Grundlagen vorstellbar ist; und sie ist ganz offensichtlich von großer Bedeutung für die Anschlussfrage, ob und in welcher Hinsicht der Wille das Prädikat „frei“ verdient. Dazu gibt es eine Reihe von Vorschlägen. Da wir den psycho-physischen Parallelismus und den Reduktionismus als ungeeignet ausgesondert haben, bleiben im wesentlichen nur noch drei Kandidaten übrig: Kausalität, Emergenz und Supervenienz. Kausalität – des Gehirns für den „Geist“ bzw. den Willen – erscheint wenig geeignet. Für dauernde und stabile Verbindungen, die zudem keine relevante Zeitdifferenz zwischen ihren Relata aufweisen, passt der Begriff der Kausalität in seiner gängigen Bedeutung schlecht. Ein naheliegendes Beispiel zur Illustration sind die durch die Gravitation vermittelten dauerhaften Wechselwirkungen zwischen den Bewegungen der Planeten. Solche Zusammenhänge würden wir eher als funktionale denn als kausale bezeichnen.<sup>147</sup> „Emergenz“ ist ein philosophischer Kunstbegriff. Die vielfachen Vorschläge zu seiner Definition laufen ungefähr auf den folgenden Kerngedanken hinaus: Emergente Phänomene (Eigenschaften, Substanzen, Ereignisse) entstehen aus fundamentaleren Entitäten, im Verhältnis zu welchen sie aber etwas prinzipiell Neues in die Welt bringen, weswegen sie von keiner denkbaren Beschreibung ihrer Entstehung auf ihre Fundamente reduziert werden können.<sup>148</sup> Man sieht sofort, dass dieser Begriff (erstens) gut zum Anliegen eines nicht-reduktiven Physikalismus passt, und dass er (zweitens) diesem Anliegen freilich nur eine neue Benennung gibt und nicht ansatzweise erkennen lässt, wie es wissenschaftlich durchzuführen wäre. Was genau Emergenz ist, wie man sich einen solchen

147 Vgl. *Keil* (Anm. 42), S. 7. Natürlich gibt es kein *Verbot*, diese Zusammenhänge dennoch unter Kausalbegriffen zu beschreiben. Damit verschöbe man aber nur das Problem: Die Frage, wie das genau vorstellbar wäre (was „Kausalität“ hier bedeuten könnte), wird dadurch weder beantwortet noch leichter.

148 S. *O'Connor/Wong*, *Emergent Properties*, in: *Stanford Encyclopedia of Philosophy* (Anm. 19); *O'Connor*, *The Metaphysics of Emergence*, in: *Noûs* 39 (2005), 658 ff.; sehr erhellend *McLaughlin* (Anm. 137).

Vorgang konstruktiv vorstellen könnte, bleibt jedenfalls für die Probleme unseres Themas genauso rätselhaft wie die Ausgangsfrage selber: die nach dem Zusammenhang von Gehirn und Geist.

Etwas Ähnliches gilt, genau besehen, auch für den Begriff der „Supervenienz“. Er besagt allgemein, dass Eigenschaften einer „höheren“ Ebene mit solchen einer fundamentaleren Ebene so zusammenhängen, dass es keine Veränderung der höherstufigen Eigenschaften ohne eine Veränderung der fundamentalen Eigenschaften geben kann. Knapp und schlagwortartig für unser Gehirn-Geist-Problem: Keine Veränderung im Mentalen ohne Veränderung im Neuronalen. „Supervenienz“ postuliert diesen Zusammenhang allerdings als asymmetrischen; umgekehrt gilt er nicht. Er schließt daher nicht aus, dass es Veränderungen im Neurophysiologischen (auf der fundamentalen Ebene) gibt, die keinerlei Änderungen im Mentalen (auf der höheren Ebene) hervorbringen. Das leuchtet ohne weiteres ein. So schwer man sich vorstellen kann, dass in einem Bewusstsein irgendeine Empfindung, ein Gedanke, ein Wille entstehen könnte, ohne dass sich irgendetwas im neuronalen Gesamtzustand des zugehörigen Gehirns veränderte, so problemlos ist die umgekehrte Vorstellung, dass in der neuronalen Textur des Gehirns Veränderungen ohne jede Folge für das zugehörige Bewusstsein auftreten können (z.B. das unbemerkte Entstehen eines Hirntumors).

Andererseits besagt der so verstandene Begriff der Supervenienz kaum mehr als eben dies: dass es eine stabile Beziehung zwischen den Eigenschaften der fundamentalen Ebene und den „supervenierenden“ Eigenschaften der höheren Ebene gibt und dass diese Beziehung asymmetrisch ist, sich in ihr also ein bestimmter *Vorrang* der fundamentalen gegenüber der höherstufigen Ebene ausdrückt. Wie genau man sich das Zustandekommen dieser Beziehung vorzustellen hat, ist damit nicht gesagt. In dieser abstrakten Bedeutung ist die Welt voller Supervenienzbeziehungen. Ein radikaler Physikalist würde sogar sagen, dass *alle* makrophysikalischen Eigenschaften des Universums über fundamentalere mikrophysikalischen „supervenieren“, und dass

die zutreffende Form der genaueren Erklärung dieses Zusammenhangs eben die der physikalistischen Reduktion ist.<sup>149</sup> Supervenienz ist also ein sehr weiter Begriff. Er ist daher in der Gehirn-Geist-Debatte ohne weiteres auch für den reduktiven Physikalisten oder den Kausalisten verwendbar. Und deshalb führt er mit einem neuen Terminus lediglich auf das alte Grundproblem zurück: Ist das Mentale (der Wille) wirklich auf die *gleiche* Weise supervenient zum Neurophysiologischen wie etwa die Makroeigenschaften eines Gegenstands, eines Gases oder einer Flüssigkeit zu den atomaren Mikroeigenschaften ihrer Bestandteile? Dagegen sprechen die oben zum Reduktionismus erörterten Argumente. Solange wir aber von einer solchen Gleichheit der Supervenienzrelationen nicht überzeugt sind, liefert der Begriff wohl ein knappes Etikett, aber keine Lösung für das Rätselhafte des Problems.

Immerhin haben wir gesehen, dass Supervenienzbeziehungen asymmetrisch sind. Darin spiegelt sich, wie ich es eben formuliert habe, ein bestimmter Vorrang der fundamentalen Ebene gegenüber der höherstufigen – für unser Problem: des Gehirns gegenüber dem Mentalen. Dieser Vorrang ist ein ontologischer. Man kann ihn sich in einer einfachen Analogie verdeutlichen. Supervenienzbeziehungen gibt es nicht nur unmittelbar in der physischen Welt, sondern auch zwischen rein begrifflich konstituierten Tatsachen: Wenn Sokrates stirbt, wird Xantippe im selben Moment zur Witwe. Wir würden nun gewiss nicht sagen, der Tod des Sokrates sei kausal für das Zur-Witwe-Werden der Xantippe; und ebensowenig würden wir beides für identisch erklären.<sup>150</sup> Das Zur-Witwe-Werden der X. superveniert vielmehr über dem Tod des S. Die Asymmetrie, um die es mir hier geht, der metaphysische (ontologische) Vorrang des Sterbens des S., drückt sich darin aus, dass wir zwar sagen können: „Indem Sokrates stirbt, wird Xantippe zur Witwe“, aber nicht umgekehrt: „Indem Xantippe zur Witwe wird, stirbt Sokrates“. In einem genau analogen Sinn ist das Neuronale

149 So etwa die eingehend begründete These von *D. Lewis* (Anm. 129).

150 Schon deshalb nicht, weil die beiden Ereignisse nicht räumlich identisch sind: Sokrates stirbt im Gefängnis, Xantippe wird da zur Witwe, wo sie sich gerade befindet (oder, wenn man will, im ganzen Universum).

vorrangig vor dem supervenienten Mentalen. Also gilt: indem bestimmte Vorgänge im Gehirn stattfinden, entsteht ein mentaler Zustand, z.B. der Entschluss zu einer Handlung. Nicht aber gilt umgekehrt: indem jemand einen Entschluss fasst, entstehen bestimmte neuronale Vorgänge in seinem Gehirn.

### 1.3 Resümee

Die Analyse des Zusammenhangs von Gehirn und Geist hat uns ein tieferes Verständnis des Problems der Willensfreiheit verschafft, nicht aber dessen Lösung. Die Frage, auf welche Weise der Wille mit seinem neurophysiologischen Fundament zusammenhängt, ist nicht weniger rätselhaft als die, ob er trotz dieses Zusammenhangs frei sein kann. Immerhin unterstreicht die Analyse, dass die Beziehung des Willens zu seinen neuronalen Grundlagen nur als eine der Supervenienz zu begreifen ist: als ontologischer Vorrang des Gehirns vor dem „Geist“.<sup>151</sup> Das beeindruckt einen kompatibilistischen Freiheitsbegriff wenig. Gezeigt hat sich aber außerdem, dass vor diesem Hintergrund auch die Möglichkeit einer „mentalen Verursachung“, also z.B. die des Auslösens und Kontrollierens von Handlungen durch einen Willen, nicht kohärent verständlich zu machen ist. Und das könnte sich für eine verantwortungsgründende Freiheit sehr wohl als destruktiv erweisen.

151 Das gilt übrigens auch für die „Veto“-Kompetenz, die Benjamin Libet dem Willen als Kontrollmöglichkeit zuschreibt, um die befürchtete freiheitsgefährdende Konsequenz seiner eigenen berühmten Forschungen zu vermeiden, wonach Handlungen stets bereits unbewusst cerebral eingeleitet sind, bevor der Entschluss dazu dem Handelnden bewusst wird (s. *Libet et al*, Time of Conscious Intention to Act..., in: *Brain* 106 [1983], 623 ff.). Dieses auch von Strafrechtlern gern zustimmend zitierte Libetsche „Veto“ ist höchst unplausibel. Wenn die Einleitung der Handlung neuronal und nicht mental erfolgt, dann wird man das Gleiche wohl auch von einem „Veto“ annehmen müssen; sonst bliebe rätselhaft, woher dieses kommen könnte. Libet bemüht sich in einem späteren Aufsatz, den Einwand zu entkräften (*ders.*, Do we have free will?, in: *Journal of Consciousness Studies* 6 [1999], 47 ff.), es gelingt ihm aber nicht. Da er keine Belege für diese Veto-Fähigkeit vorweisen kann, zieht er sich zuletzt auf das Argument zurück, es gebe auch „keine experimentellen Beweise“ *dagegen* (ebda., S. 53). Das ist freilich bei weitem zu wenig.

## 2. *Andershandelnkönnen (PAM) als notwendige Bedingung für Freiheit und Verantwortlichkeit? – Der Kompatibilismus Harry G. Frankfurt*

Bislang unberührt geblieben ist unsere anfangs formulierte Grundthese PAM: dass eine Handlung (bzw. der Entschluss dazu) nur dann „frei“ genannt werden kann, wenn der Täter auch anders hätte handeln oder jedes Handeln hätte unterlassen können (oben sub II.3.). Die Ergebnisse unserer bisherigen Untersuchung legen es aber nahe, die Möglichkeit einer Erfüllung von PAM mindestens für zweifelhaft zu halten. Wenn der Wille über neuronalen Prozessen im Gehirn superveniert, die ihrerseits naturgegebenen Regularien folgen und somit keiner menschlichen Kontrolle zugänglich sind, dann erscheint der Schluss, dass kein Handelnder jemals anders handeln konnte, als er gehandelt hat, prima facie unausweichlich. Das ändert, wie wir gesehen haben, selbstverständlich nichts daran, dass die meisten Handelnden vor vielen ihrer Entscheidungen überlegen, Alternativen prüfen, abwägen und schließlich aus wohlwogenen Gründen eine davon wählen können, so wenig wie es ausschließt, dass sie dies alles, kantianisch gesprochen, nur „unter der Idee der Freiheit“ tun können, also mit dem subjektiven Empfinden, sie selber seien die Urheber ihrer jeweiligen Entscheidung und nichts und niemand zwingt sie dazu. Denn alles dies, die gesamte Kette des Raisonnements mitsamt dem begleitenden subjektiven Erleben der eigenen Entscheidungsfreiheit, kann gleichwohl lückenlos neuronal determiniert sein. Und dass es dies tatsächlich ist, dafür sprechen neben den obigen philosophischen Analysen sämtliche verfügbaren empirischen Befunde weitaus nachdrücklicher als für irgendeine andere Möglichkeit.<sup>152</sup> Wenn sich das aber so verhält, dann ist in einem präzisen Sinne für keinen Handelnden, mag er moralisch oder rechtlich falsch oder richtig handeln, ein Andershandelnkönnen möglich.

152 S. speziell zum Thema Willensfreiheit nur *Spence*, *Free Will in the Light of Neuropsychiatry*, in: *Philosophy, Psychiatry & Psychology* 1996, 75 ff., mit Anm. von *Frith, Libet, Stephens* und Erwiderung *Spence*.

## 2.1 Frankfurts Angriff auf PAM

Aber ist *das* wirklich eine Voraussetzung für Freiheit und Verantwortlichkeit? Der amerikanische Philosoph Harry G. Frankfurt hat dies vor fast vierzig Jahren mit einem berühmt gewordenen Gedankenexperiment zu widerlegen versucht. Ich übernehme hier nur dessen abstrakte Struktur und fülle sie mit einem eigenen, etwas plastischeren Fallbeispiel aus<sup>153</sup>:

A rollt mit seinem PKW eine steile Straße hinab, den rechten Fuß neben dem Gaspedal auf dem Boden, als ihm betrunken sein Feind B vors Auto läuft. A hält das Lenkrad und seine Füße unverändert und überfährt, wie beabsichtigt, B tödlich. Was A nicht weiß: X, der jedenfalls sicherstellen will, dass A den B überfährt, und der zuvor A's Auto entsprechend präpariert hat, beobachtet das Geschehen über einen Monitor und hält sich bereit, per installierter Fernsteuerung Bremsen und Lenkung des PKW des A sofort zu blockieren, sollte dieser sich anschicken zu bremsen oder um B herumzulenken. A kommt jedoch gar nicht auf diese Idee. Er will B tödlich überfahren und tut das auch. Er könnte jedoch auch dann nichts anderes tun, wenn er das wollte und versuchte; denn X würde es verhindern. Freilich muss X nicht eingreifen; denn A tut von sich aus, was X will.

Es stehe außer Frage, so Frankfurt, dass A für sein Handeln genauso verantwortlich sei, wie er es ohne Überwachung durch X wäre. X habe ja keinerlei Beitrag zu dem Geschehen geleistet.<sup>154</sup> Gleichwohl hätte A wegen der Kontrolle durch X nicht anders handeln, sein konkretes Tötungshandeln nicht vermeiden können. Der Umstand dieser Überwachung sei zwar eine hinreichende Bedingung dafür, dass A so

153 S. Frankfurt, *Alternate Possibilities and Moral Responsibility*, in: *Journal of Philosophy* 66 (1969), 829 ff. – Frankfurts Gedankenexperiment hat in der anschließenden Diskussion viele Autoren zur Erfindung ähnlicher Beispiele angeregt (sog. „Frankfurt-Type Examples“). Die Literatur dazu ist inzwischen unüberschaubar; instruktiv *Betzler/Guckes* (Hrsg.), *Autonomes Handeln. Beiträge zur Philosophie von Harry G. Frankfurt*, 2000.

154 Strafrechtlich gesprochen: X hat nicht einmal zum Versuch eines Totschlags (in mittelbarer Täterschaft) „unmittelbar angesetzt“ (§ 22 StGB), sich also keines Tötungsdelikts schuldig gemacht.

handeln musste, wie er gehandelt hat; doch sei er dafür keine notwendige Bedingung. In einer vollständigen Erklärung, *warum* A so gehandelt hat, komme der Umstand der Kontrolle durch X nicht vor. Dass A nicht anders hätte handeln können, sei daher irrelevant. Denn auch wenn er's gekonnt hätte, hätte er nicht anders gehandelt. Also müsse PAM aufgegeben werden.<sup>155</sup>

Beispiel und Erläuterung haben zunächst eine hohe Suggestivität. Sieht man schärfer hin, stellen sich aber Zweifel ein. Die Frage, ob A anders hätte handeln können, hängt davon ab, wie man sein Handeln beschreibt. Tut man es so: „A hat B bewusst tödlich überfahren“, dann hätte A nicht anders handeln können. Verneint man aber die Vollständigkeit dieser Beschreibung und erweitert sie daher um ein voluntatives Handlungselement, dann sieht die Sache anders aus: „A hat B bewusst und gewollt tödlich überfahren“. *Dazu* hatte A offensichtlich eine Alternative, nämlich die, den B zwar bewusst, aber unwillentlich zu überfahren.

Nun mag man bestreiten, dass dies für das Handeln des A einen Unterschied ausmache, sei es, weil man (wenig plausibel) die subjektive Seite überhaupt nicht zur Handlung rechnen will, sei es, weil man dafür beim Handelnden schon die *Kenntnis* dessen, was er objektiv tut, ausreichen lässt und ein begleitendes voluntatives Element für entbehrlich hält.<sup>156</sup> Doch können wir die Frage, ob die Differenz eines (und ggf. welches) subjektiven Elements aus zwei exakt gleichen äußeren Verhaltensweisen zwei verschiedene Handlungen macht, hier offenlassen. Denn ein zweiter genauerer Blick zeigt, dass sie in unserem Fall sogar einen Unterschied ums Ganze bedeutet: den zwischen Handeln und Nichthandeln. Hätte A pflichtgemäß versucht zu bremsen oder auszuweichen, so hätte er festgestellt, dass ihm das unmöglich war. Zwar hätte er B dann auf genau dieselbe Weise tödlich überfahren, doch wäre ihm dieses Weiterfahren nicht einmal mehr als Handlung, sein Nichtbremsen etc. nicht als Unterlassen zuzurechnen

155 *Frankfurt* (Anm. 153), 836 f.

156 Um diesen letzteren Punkt wird in der strafrechtlichen Vorsatzlehre bekanntlich seit eh und je gestritten.

gewesen. Es wäre aus seiner Sicht ein böses Schicksal, aus objektiver Sicht allein die böse Tat eines anderen gewesen. A hätte also durch sein Anderswollen und den Versuch, dieses handelnd umzusetzen, sich zu einem Nichthandelnden gemacht. Ein Nichthandeln ist aber im Vergleich zum Handeln ein Andershandeln, und zwar auch dann, wenn beide äußerlich völlig identisch sind (wie gerade unser Fall zeigt).

Man kann das Beispiel übrigens problemlos so modifizieren, dass *ceteris paribus* nicht ein kontrollierender X, sondern ein natürlicher Umstand, etwa ein Ausfall der Elektronik in A's PKW, jede Möglichkeit des Bremsens oder Lenkens ausgeschlossen hätte, aber ebenfalls ohne Relevanz geblieben wäre, weil A gar nicht daran dachte, zu bremsen oder zu lenken. Nach Frankfurt ändert dies für die Zurechnungs- und die Schuldfrage nichts. Das ist freilich zweifelhaft. Im Strafrecht würde in dieser Variante jedenfalls nur wegen Versuchs bestraft; denn A hätte hinsichtlich seines *Tötungshandelns* tatsächlich nicht „anders handeln“, nämlich den Tod unter keinen Umständen vermeiden können. Also wird ihm dieser nicht zugerechnet. Doch hält das Strafrecht, wie Frankfurt, daran fest, dass jedenfalls ein „böser Wille“ im Spiel gewesen ist, der bei A mit der Zurechnung eines Versuchs ohne weiteres als „frei“ vorausgesetzt würde.<sup>157</sup>

Frankfurts These, ein Andershandelnkönnen sei für die Zurechnung von Verantwortlichkeit nicht erforderlich, wird also von der bislang erörterten Variante seines Beispiels nicht plausibel gemacht. Denn hier *kann* A eben anders handeln, und das unterscheidende Kriterium ist sein Wille. Aber das ist zugleich der Grund, warum *dieses* Frankfurt-Szenario unser eigentliches Problem noch gar nicht berührt: Hätte A denn auch anders wollen können?

157 Man fragt sich vielleicht, warum dann in der ersten Variante (im Hintergrund kontrollierender X), nicht auch nur wegen Versuchs bestraft würde, da dort der A den Tod genauso wenig hätte vermeiden können. Das ist richtig, und manche Strafrechtler würden genau deshalb hier ebenfalls nur wegen Versuchs bestrafen. Die zutreffende Lösung muss aber ausschließen, dass sich zwei Schurken wechselseitig durch Verweis auf den jeweils anderen rechtlich entlasten (A hätte den Erfolg nicht vermeiden können, weil X bereitstand, dieser hat gar nicht gehandelt, weil das jener schon hat). Die Frage ist im Strafrecht freilich noch wenig erörtert.

Doch lässt sich die Kontrollinstanz in Frankfurts Gedankenexperiment unschwer sozusagen bis ins Gehirn des handelnden A zurück verlegen. Dann geriete nicht nur A's Handeln, sondern schon sein Wille unter fremde Kontrolle. X würde etwa über feinste, von ihm kontrollierte Sensoren im Schädel des A die Vorgänge in dessen Gehirn genau beobachten, im Falle des Aufbaus eines sog. „Bereitschaftspotentials“ zum Bremsen oder Herumlenken sofort korrigierend eingreifen und ein gegenläufiges Bereitschaftspotential (eines zum Überfahren) erzeugen.<sup>158</sup> Und dies geschähe, bevor A selbst auch nur das mindeste von dem bewusst werden könnte, was die beginnenden neuronalen Vorgänge in seinem Kopf (Bereitschaftspotential zum Bremsen) soeben an Wollen und Handeln bei ihm auszulösen im Begriff wären, wenn X nicht intervenieren würde. Erneut soll aber gelten: X muss gar nicht intervenieren. A baut sozusagen von Anfang an das gewünschte „böartige Bereitschaftspotential“ selber auf.

Nun erst erhält Frankfurts Beispiel die erforderliche Schärfe.<sup>159</sup> Denn jetzt hätte A buchstäblich auch nicht anders „wollen“, sich nicht anders entscheiden können. Und dennoch würden wir ihn ohne weiteres für verantwortlich halten, strafrechtlich gesprochen: zumindest als Versuchstäter. Er hätte, so könnte man sagen, eben mit dem Aufbau eines anderen („guten“) Bereitschaftspotentials beginnen sollen – wiewohl ihm das, hätte er's getan, wegen der dann unvermeidlichen Intervention des X nicht einmal hätte bewusst werden können. Da er sich aber von Anfang für das Böse entschieden hat, wird er mit Recht getadelt und bestraft. Belegt dies Frankfurts These, dass ein Andershandelnkönnen

158 Zu Fragen dieses (berühmten) „Bereitschaftspotentials“ *Libet, Time of Conscious Intention* (Anm. 151); *ders.*, The neural time factor in conscious and unconscious events, in: Ciba Foundation Symposium 174: Experimental and Theoretical Studies of Consciousness, 1993, S. 123 ff.; eingehend zu Libets Experiment *Detlefsen* (Anm. 10), S. 278 ff.

159 Frankfurt sieht und erwähnt (etwas undeutlich) die Möglichkeit der Erstreckung seines Gedankenexperiments auf die Willenskontrolle; s. *Frankfurt* (Anm. 153), 836 Fn. 3. Diese Flexibilität von Frankfurts Szenario unterschätzt *J. Nida-Rümelin* (Anm. 24), S. 95 ff., der eine recht grobe Variante skizziert und ablehnt, die ein rein äußeres (vermeintliches) Nichtandershandelnkönnen betrifft und die sich daher auch leicht ablehnen lässt.

keine Voraussetzung für Verantwortlichkeit ist, weil nicht einmal ein Anderswollenkönnen dafür zwingend erforderlich ist?

Das hängt davon ab, aus welchem Grund ein Anderswollen unmöglich ist. Liegt der Grund darin, dass jeder potentielle andere Wille von einer externen Kontrollinstanz *verhindert* würde, sollte er sich denn sozusagen anschicken zu entstehen, so schließt das weder Freiheit noch Verantwortlichkeit aus, wenn nur der tatsächliche Wille des Handelnden allein aus ihm selbst stammt. (Das gilt selbst dann, wenn das hypothetische Entstehen eines anderen Willens schon im Keim, nämlich noch unterhalb der Bewusstseinsschwelle, erstickt würde.) Ist dagegen ein Anderswollen deshalb unmöglich, weil eine externe Instanz genau diesen tatsächlichen Willen im Handelnden *erzeugt*, etwa durch Hypnose oder durch direkte Gehirnmanipulation, und deshalb kein anderer (genuin eigener) Wille des Handelnden entstehen kann, so würde niemand einen solchen fremderzeugten Willen frei nennen. Knapp und technischer: Jede *negative* Freiheit (zur Willensalternative) kann fehlen, wenn nur die *positive* Freiheit (zur Willenserzeugung) gegeben war. Das reicht als Grundlage für Verantwortlichkeit aus.

Damit macht unsere Zuspitzung des Frankfurtschen Gedankenexperiments die Grenzen seiner Beweiskraft ganz deutlich. Denn zu der Frage, ob ein alternativer Wille immerhin positiv hätte initiiert werden können, sagt das Beispiel gar nichts. Es stellt nur fest, dass die potentielle externe Blockade einer Willensalternative Verantwortlichkeit nicht ausschließt, wenn sich der Handelnde schon intern und ganz von selbst gegen jede solche Alternative entschieden hat. Damit stehen wir aber erneut vor unserer Ausgangsfrage: Hätte ein Handelnder *positiv* einen anderen Willen bilden (oder doch mit dieser alternativen Willensbildung beginnen) können? *Dieses* Anderswollenkönnen scheint unser Begriff von Freiheit und Verantwortlichkeit weiterhin zu fordern. Und dazu sagt Frankfurts Szenario nichts. Es fragt nicht danach.<sup>160</sup>

160 Aus libertär-inkompatibilistischer Sicht sind Einwände formuliert worden, die, wiewohl in eine andere Richtung zielend, auf einer ähnlichen Erwägung beruhen: Falls die allererste (ggf. unbewusste) Initiation eines bösen Willens bei A determi-

## 2.2 Autonome Selbstvergewisserung statt Andershandelnkönnen?

Allerdings entwickelt Frankfurt in späteren Aufsätzen auch auf diese Frage eine Antwort. Deren Prämisse ist zunächst eine Einsicht, die sein Gedankenexperiment, unbeschadet anderweitiger Grenzen, doch immerhin nahelegt: Die Frage, ob ein konkreter Handlungsentschluss jemandes *eigener* Entschluss ist, ist für die Frage der Verantwortlichkeit bedeutsamer als die, ob der Handelnde diesen Entschluss hätte vermeiden, sich also anders hätte entschließen können. In gängiger Terminologie: die autonome Urheberschaft an einem Willensentschluss ist wichtiger als das Prinzip alternativer Möglichkeiten.<sup>161</sup> Hier knüpft Frankfurt nun an und schlägt Kriterien dafür vor, unter welchen Bedingungen ein solcher Willensakt als jemandes eigener Wille gelten kann. Das sei dann der Fall, wenn ein primärer (determinierter oder indeterminierter) Willensimpuls in einem zweiten Akt der bewussten inneren Annahme als eigenes Wollen akzeptiert und damit gleichsam ratifiziert werde. Frankfurt spricht von „first-order volition“ für den noch ungefilterten ursprünglichen Impuls, der dann von einer „second-order volition“, einem reflektierten inneren Akt zweiter Stufe, beglaubigt werde. Das sei der entscheidende Akt innerer Aneignung. Personale Autonomie sei also ein hierarchisch gestufter Prozess der Selbstvergewisserung im Wollen und Entscheiden. Ein dergestalt ratifizierter Wille reiche zur Begründung von Verantwortlichkeit aus.

niert sei, dann sei das Szenario für Libertarier witzlos, weil sie diese Prämisse ablehnten; sei sie aber indeterminiert, dann *könnte* A anders entscheiden oder doch eine andere Entscheidung einleiten (auch wenn X dann sofort intervenieren würde); so etwa *Widerker*, *Libertarianism and Frankfurt's Attack on the Principle of Alternative Possibilities*, in: *Philosophical Review* 104 (1995), 247 ff.

- 161 Zwei Anmerkungen: (1.) Beides ist nicht dasselbe (wie oft unbesehen angenommen wird), eben weil das potentielle Andershandeln/Anderswollen durch Dritte ausgeschlossen, d.h. *potentiell* unterbunden werden kann, wovon aber das autonome innere Hervorbringen des Willens nicht berührt wird. (2.) „Autonomie“ ist hier nicht identisch mit dem gleichlautenden Begriff bei Kant, der Autonomie nur dann bejaht, wenn sich die eigene Maximenbildung dem Kategorischen Imperativ unterordnet.

Weder die Möglichkeit eines Anderswollens noch ein Indeterminiertsein dieses Willens sei dafür erforderlich.<sup>162</sup>

Das ist nur eine grobe Skizze. Frankfurt hat die Bestimmung dessen, was „second-order volitions“ enthalten müssen, um als Beglaubigungsinstanz für Autonomie tauglich zu sein, erheblich genauer formuliert, später auch mehrfach verändert und insbesondere um (freilich wenige) normativierende Elemente erweitert. Ob das eine überzeugende Konzeption von Autonomie ist, mag hier offen bleiben.<sup>163</sup> Für uns ist etwas anderes bedeutsam: Kann es richtig sein, neben einem solchen Modell autonomer Selbstvergewisserung die Frage, ob die relevanten „Volitionen“ auf den beiden Stufen der Willenserzeugung und -aneignung *determiniert* zustandekommen oder nicht, gänzlich zu ignorieren, ihr keine Bedeutung für die Frage von Freiheit und Verantwortlichkeit mehr beizumessen?

Ich glaube das nicht. Das Problematische von Frankfurts Auffassung wird deutlich, wenn man – per Science Fiction – einen Fall unterstellt, in dem jene „Volitionen“ (ggf.) nicht durch Naturprozesse, v.a. die neuronalen Funktionen des Gehirns, sondern durch die steuernde Willkür einer dritten Person determiniert werden. Nehmen wir also an, der Neurochirurg N programmiert anlässlich einer Hirntumoroperation an P dessen für „Volitionen“ zuständige Hirnareale im limbischen System auf eine Weise um, dass P, der bisher keinerlei Neigung zur Kunst gezeigt hat, vielmehr passionierter Sportler gewesen ist, nun umgekehrt zum enthusiastischen Kunstsammler wird und den Sport aufgibt, weil er an ihm jedes Interesse verloren hat.<sup>164</sup> P hat zu dieser Umprogrammierung seines Seelenlebens keine Einwilligung gegeben; er identifiziert sich aber nun, so wollen wir annehmen, begeistert mit

162 S. v.a. *Frankfurt*, Freedom of the Will and the Concept of a Person, in: Journal of Philosophy 68 (1971), 5 ff.

163 Auch hierzu gibt es eine Flut an Debattenbeiträgen; s. statt aller die Aufsätze in *Betzler/Guckes* (Anm. 153), sowie *Ekstrom*, Alienation, Autonomy and the Self, in: Midwest Studies in Philosophy 29 (2005), 45 ff.

164 Das Irreale des Szenarios ist kein Einwand gegen seine Tauglichkeit zur Veranschaulichung. Gedankenexperimente sollen nicht die Wirklichkeit abbilden, sondern die Struktur von Argumenten transparent machen.

seiner neuen Leidenschaft und kann seine frühere gar nicht mehr begreifen. Nach Frankfurt müssten wir die Hinwendung des P zur Kunst als Resultat einer autonomen Entscheidung beurteilen. Das leuchtet nicht ein. Die relative Harmlosigkeit des gewählten Beispiels mag das zunächst verdunkeln: Erstens ist man vielleicht geneigt, den Tausch Kunst gegen Sport eher vorteilhaft zu finden, zumal P das (jedenfalls nachträglich) selbst tut; und zweitens stellt sich allein wegen der Hinwendung zu einer neuen und honorigen Leidenschaft normalerweise keine Frage von Verantwortlichkeit. Man muss aber das Beispiel nur ins rundum Negative wenden: N sei ein Dr. Frankenstein, der den vor der Operation freundlichen und liebenswürdigen P durch eine gezielte Gehirnmanipulation zu einem miserablen Zeitgenossen macht, der Moral- wie Rechtsnormen nach Belieben verletzt. P identifiziert sich aber, und zwar genau deshalb, weil er so „programmiert“ wurde, in jeweiligen „second-order volitions“ mit jedem einzelnen seiner tadelnswerten Handlungsentschlüsse. Soll ihn das verantwortlich machen? Das wäre schwerlich akzeptabel.

### 2.3 Resümee; Überleitung zum Schuldprinzip

Und damit sind wir zurück bei unserem Ausgangsproblem: der Frage nach dem Zusammenhang von „neuronaler Determination und Schuld“. Denn was genau ist *aus der Sicht eines Handelnden* eigentlich der relevante Unterschied zwischen einer neuronalen Programmierung seiner Bösartigkeit durch einen Dr. Frankenstein und einer durch die Natur? Nach unseren obigen Überlegungen zur Supervenienz des Mentalen spricht alles dafür, dass die handlungswirksame „Unordnung“ in den willenserzeugenden Hirnarealen bei einem „natürlichen“ Schurken nicht *prinzipiell* anders aussieht als bei einem künstlich programmierten.<sup>165</sup> Und mit dieser Erwägung führt unser

165 Vielerlei Einzelheiten mögen selbstverständlich variieren, nicht anders als zwischen zwei „natürlichen“ Schurken mit gleichartigen Charakterdefekten auch. Das betrifft die vieldiskutierte These (ursprünglich *H. Putnams*) von der „multiplen Realisierbarkeit“ mentaler Eigenschaften. Hier kann das unerörtert bleiben. Gemeint ist nur, dass bei gleichartigen mentalen Defekten im wesentlichen die gleichen Hirnareale betroffen

Gedankenexperiment, so abstrus es auf den ersten Blick anmuten mag, unversehens zurück ins wirkliche Leben. Man erinnere sich an den oben in Anm. 93 skizzierten Fall des Bahnarbeiters Phineas Gage aus dem 19. Jahrhundert. Dort war es zwar kein Science-Fiction-Chirurg, wohl aber eine schwere Eisenstange von 3 cm Durchmesser, die Schädel und Frontalhirn von Gage durchdrang und dessen neuronale „Umprogrammierung“ herbeiführte – von einem vorher freundlichen und allseits beliebten zu einem miserablen, schwer erträglichen Zeitgenossen.<sup>166</sup> Hier ist ein aktuelles Beispiel:

L, ein 40-jähriger Lehrer in den USA, verheiratet, Familienvater und bisher ohne jeden Konflikt mit dem Recht, entwickelte im Jahr 2000 starke pädophile Neigungen. Er begann, kinderpornographisches Material in großen Mengen zu sammeln, stellte seiner minderjährigen Stieftochter zunehmend sexuell nach, versuchte sein Verhalten freilich vor seiner Frau zu verbergen, weil er es selber inakzeptabel fand, konnte aber seine Neigungen trotz entsprechender Anstrengungen nicht beherrschen. Sein Tun wurde offenkundig, er wurde wegen sexueller Belästigung von Kindern angeklagt und verurteilt; statt der Haftstrafe konnte er ein Therapieprogramm für Pädophile wählen. Dies tat er, wurde aber bald ausgeschlossen, weil er auch dort seine sexuellen Neigungen nicht zu zügeln wusste. Am Tag vor Antritt der Haftstrafe ging L wegen starker Kopfschmerzen in eine Klinik. Die neurologische Untersuchung zeigte einen hühnereigroßen Hirntumor im orbitofrontalen Cortex, in einem Hirnareal, dessen Funktion als Zentrum der Impulskontrolle für Aggressionen und anderes normverletzendes Verhalten vielerlei Art wissenschaftlich bereits bekannt war. Der Tumor wurde operativ entfernt, die pädophilen Neigungen verschwanden. Zehn Monate später litt L erneut unter ständigen Kopfschmerzen und begann wieder, heimlich kinderpornographisches Material zu sammeln. Eine Magnet-Resonanz-Tomographie im Okto-

sein dürften. Die Ähnlichkeit ist also eine funktionale, nicht unbedingt eine biologische.

166 S. Anm. 93; auch <http://www.deakin.edu.au/hmnbs/psychology/gagepage/>, mit dokumentarischem Material.

ber 2001 ergab, dass der Tumor sich neu gebildet hatte. Er wurde erneut entfernt, die pädophilen Neigungen verschwanden sofort wieder.<sup>167</sup>

Es stand außer Zweifel, dass Auslöser der Pädophilie des L der Hirntumor war. Das muss man aber genauer erwägen. Denn genauso zweifelsfrei war, dass die pädophilen Impulse nicht unmittelbar aus der unstrukturierten Tumormasse selbst stammten. Vielmehr entstanden sie aus den Läsionen, aus der Unordnung, die der Tumor im angrenzenden Hirngewebe des präfrontalen Cortex verursachte. Das legt die Frage nahe, was eigentlich *diese* Unordnung von der eines anderen, tumorfreien, aber ebenfalls pädophile Neigungen erzeugenden Gehirns substantiell bzw. prinzipiell unterscheidet. Die weitaus plausibelste Antwort lautet: nichts. Die allermeisten Strafgerichte, bei uns nicht anders als in den USA, dürften freilich im Fall des L weit eher geneigt sein, die Frage der Schuldunfähigkeit zu bejahen als im „Normalfall“ eines Pädophilen ohne sichtbare Läsionen im Gehirn. Was genau rechtfertigt – vom Norminhalt des § 20 StGB vorläufig abgesehen – eine solche Ungleichbehandlung?

Man könnte antworten, die Betroffenen würden ja, jedenfalls de lege, in entscheidender Hinsicht gar nicht ungleich behandelt. Denn es komme nicht auf die An- oder Abwesenheit des Tumors, sondern nur auf die Frage an, ob jemand – mit oder ohne Tumor – in der Lage (gewesen) sei, seine pädophilen Handlungsimpulse zu *beherrschen*, egal aus welchem Grund sie in seinem Gehirn entstanden sein mögen. Die Auskunft entspricht, wie ein Blick in § 20 StGB zeigt, dem geltenden Recht.<sup>168</sup> Doch legt sie sofort die Anschlussfrage nahe, wie ein

167 Genauer Bericht des Sachverhalts und der Befunde bei *Burns/Swerdlow*, Right Orbitofrontal Tumor With Pedophilia Symptom and Constructional Apraxia Sign, in: *Archives of Neurology* 60 (2003), 437 ff.

168 Gleichwohl kämen die meisten deutschen Gerichte im Fall des L mit hoher Wahrscheinlichkeit *de facto* zur Annahme der Schuldunfähigkeit, dagegen in vermutlich keinem oder kaum einem anderen Fall von Pädophilie, in dem außer der abweichenden Sexualneigung nicht auch noch eine (weitere!) manifeste psychische Erkrankung vorläge. In Fällen „gewöhnlicher“ Kriminalität *ohne* eine der in § 20 genannten Psychopathien wird nach der Beherrschbarkeit des kriminellen Antriebs gar nicht gefragt; sie wird vielmehr vorausgesetzt.

Betroffener denn seine Beherrschungsimpulse erzeugen und ihnen das Übergewicht über seine Handlungsimpulse soll sichern können, wenn solche Kontrollimpulse doch *ganz genauso* wie die Antriebsimpulse in seinem Gehirn entstehen, also natürlichen Determinanten unterliegen? Der Tumor des L hat nicht nur dessen pädophile Neigungen erzeugt, sondern zugleich seine Kontrollfähigkeiten reduziert. Die oben formulierte Frage, was denn die pädophilieerzeugende Unordnung im Tumorgehirn von der pädophilieerzeugenden im tumorfreien Gehirn prinzipiell unterscheiden soll, stellt sich mit Blick auf die inhibierenden Kontrollzentren des Gehirns ganz genauso: Was sollte denn die tumorbedingte Funktionsschwäche der neuronalen Kontrollinstanzen von der nicht tumorbedingten prinzipiell unterscheiden?

Nicht wenige Strafrechtler würden wohl folgende Antwort geben: Der für eine Schuldfähigkeit bedeutsame Unterschied liege a limine nicht im Gehirn, sondern in den unterschiedlichen normtreuen *Anstrengungen*, zu denen der freie Wille der Betroffenen jeweils noch imstande sei.<sup>169</sup> Aber das dreht gewissermaßen das Rad unserer ganzen bisherigen Untersuchung auf einen allzu unbedarften Ausgangspunkt zurück. Selbstverständlich kann man sich anstrengen, Handlungsimpulse zu unterdrücken, und nicht selten mit Erfolg. Aber wo kämen denn solche Anstrengungen her, wenn nicht genau von dort, wo sie (nach neurobiologischer Auskunft) *in all ihrer Stärke und Schwäche* ihren Produktionsort haben: aus dem Gehirn? Keine der verfügbaren empirischen Evidenzen deutet darauf hin, dass die Annahme, man könne neuronal erzeugte Handlungsimpulse über sozusagen präter-neuronale, „frei gewollte“ Anstrengungen korrigieren, auch nur einen plausiblen Sinn haben, geschweige denn zutreffen könnte.

Einen Kompatibilisten muss all das nicht irritieren. Die Mängel der Frankfurtschen Konzeption lassen sich durch normativierende Ergänzungen beheben, vor allem durch den Ausschluss extern erzeugter

169 Exemplarisch *Wessels/Beulke*, Strafrecht Allgemeiner Teil, 36. Aufl. 2006, Rn. 400: „Die innere Berechtigung des Schuldvorwurfs“ liege in der Möglichkeit des Menschen „bei Anspannung seines ‚Rechtsgewissens‘ [...] das *rechtlich Verbotene zu vermeiden*“ (sinngemäßes Zitat aus der berühmten Entscheidung BGHSt 2, 194, 200).

„second-order volitions“ aus dem Bereich echter Autonomie. Dann bleibt der Grundgedanke jedes Kompatibilismus übrig. Knapp und salopp: Dass meine Neigungen, Charakterzüge, Fähigkeiten und Präferenzen genauso wie meine darauf beruhenden Handlungsentschlüsse aus dem determinierten Funktionieren meines Gehirns stammen, ist für eine wohlverstandene Freiheit – nämlich eine, die es wert ist, gewollt zu werden<sup>170</sup> – nicht nur unschädlich; es ist vielmehr geradezu Voraussetzung dafür, dass mein Wille wirklich mein authentisch-eigener ist. (Wessen Wille wäre denn ein indeterminierter?) In einem präzisen Sinn *bin* ich einfach all das, was mein Gehirn als mein mentales Selbst erzeugt; die Unterscheidung zwischen „mir“ und meinem Gehirn ist daher von Anfang an verfehlt.<sup>171</sup> Unfrei sind nur Willensentschlüsse, die ich selbst als (wodurch immer) erzwungen erlebe, und solche, die eindeutig persönlichkeitsfremd sind, also aus dem Rahmen meines empirisch ermittelbaren mentalen Normal-Ich weit herausfallen. Mehr ist nicht zu haben, aber auch nicht zu wünschen.<sup>172</sup>

Das ist prima facie ein attraktiver Freiheitsbegriff – solange, wie mit dem gehirngelundenen Selbst des eigenen Charakters und seiner Entscheidungen alles halbwegs gut geht, wie sich also alles im Rahmen einer im großen und ganzen normtreuen Lebensführung hält und das jeweilige Selbst sich nicht als das eines kriminellen „Hangtäters“ materialisiert. Ist freilich das letztere der Fall, dann stellt sich auch für den Kompatibilisten unabweisbar die Frage, ob man für das Sosein seines Gehirns verantwortlich sein kann.<sup>173</sup> Das erscheint in hohem

170 Formulierung nach *Dennett*, *Elbow Room: The Varieties of Free Will Worth Wanting*, 1984.

171 Das heißt keineswegs, dass Personen als ganze mit ihren Gehirnen identisch seien! Es bezieht sich nur (und auch dies mit Einschränkungen) auf das entscheidende und handelnde Ich.

172 So verstehe ich die Positionen von *Pauen*, *Illusion Freiheit?*, 2004, und (mit etwas anderen Nuancierungen) von *Beckermann*, *Freier Wille – Alles Illusion?*, in *Barton* (Hrsg.), „...weil er für die Allgemeinheit gefährlich ist!“, Prognosegutachten, Neurobiologie, Sicherungsverwahrung, 2006, S. 293 ff.; ähnlich *Gomes*, *Volition and the Readiness Potential*, in: *Journal of Consciousness Studies* 6 (1999), 59 ff.

173 Dezidiert und scharfsinnig dagegen *G. Strawson*, *The Impossibility of Moral Responsibility*, in: *Philosophical Studies* 75 (1994), 5 ff.; *ders.*, *The Bounds of Freedom*, in: *Kane* (Anm. 16), S. 441 ff.

Grade kontraintuitiv. Und daran wird deutlich, dass sich für den Kompatibilismus die Begriffe Freiheit und Verantwortlichkeit voneinander lösen, weit stärker jedenfalls als für den Inkompatibilismus, und wohl auch stärker, als vielen Kompatibilisten bewusst ist. Es ist eine Sache zu sagen: „Genau so sein und handeln zu können, wie man als Persönlichkeit durch die Entwicklung des eigenen Gehirns geworden ist, das macht die wünschenswerte Freiheit aus“, und eine ganz andere, für dieses Sein und Handeln verantwortlich gemacht zu werden, wenn die gehirngedundene Entwicklung der Persönlichkeit schief, nämlich in die Bahnen einer kriminellen Karriere gelaufen ist. Kurz: kompatibilistische Freiheitsbegriffe sind zwar, so meine ich, weitaus vernünftiger begründet als inkompatibilistische. Sie bieten aber der Zurechnung von Verantwortlichkeit, Schuld und Strafe nur eine brüchige Grundlage. Im übrigen entspricht die kompatibilistische Ausgrenzung krass persönlichkeitsfremder Handlungen als „unfrei“ weder den Zurechnungsregeln unserer Alltagsmoral, noch (und schon gar nicht) denen unseres Strafrechts. Eine Straftat mag sich vom Charakter und der bisherigen Lebensführung des Täters als noch so unerklärliches Rätsel abheben: weist er nicht zugleich (und *außerhalb* der Tatbegehung) einen manifesten psychischen Defekt auf, so wird nach seiner Schuldfähigkeit nicht einmal gefragt, geschweige denn würde sie wegen Unfreiheit verneint. Und ist ein solcher Defekt nachweisbar, so begründet allenfalls er, nicht aber die Persönlichkeitsfremdheit der Tat, die Schuldunfähigkeit des Täters.

Und damit stehen wir unmittelbar vor der Frage nach Grund, Legitimation und Reichweite des strafrechtlichen Schuldprinzips. Ihr will ich mich im letzten Abschnitt zuwenden.