

Elemente der Philosophie des Geistes und der Wissenschaftstheorie (2)¹

Reduzierbarkeit und freier Wille

U. Mortensen

¹KognitionFreierWilleCd.pdf, 11. 02. 2025, letzte Fassung 05. 09. 2025

Inhaltsverzeichnis

1	Zur Philosophie des Freien Willens	3
1.1	Freier Wille, Moral, und die Wissenschaft	3
1.1.1	Physik, Kausalität, und freier Wille	7
1.1.2	Libets Experimente, Bereitschaftspotential	15
1.2	Einige philosophische Ansätze	20
1.2.1	Immanuel Kant und David Hume	20
1.2.2	Martin Heisenbergs Befunde	26
1.2.3	Enaktivismus und freier Wille	28
1.2.3.1	Frederik I. Dretske	29
1.2.3.2	Jürgen Habermas	30
1.2.3.3	Julian Nida-Rümelin	41
1.2.3.4	Alva Noë	45
1.2.4	Max Plancks Betrachtungen	49
1.2.5	Seth Lloyds rekursive Argumentation	51
1.2.6	Freier Wille und moralische Verantwortung	54
1.2.7	Galen Strawsons Folgerungen	57
1.2.8	Albert Einsteins Bekenntnis	58
1.3	Philosophie und Empirie	58
2	Anhang	70
2.1	Turingmaschinen	70
2.2	Turings Halte-Problem	71
2.3	Lloyds Beweis	73
	Literatur	75
	Index	84

1 Zur Philosophie des Freien Willens

1.1 Freier Wille, Moral, und die Wissenschaft

Spätestens seit Parmenides (520/515 – 460/455 v.Chr.) scheiden sich im westlichen Kulturkreis die Geister an der Frage, ob der menschliche Wille frei ist oder nicht. Das Frage nach der Willensfreiheit ist weniger ein Thema der Empirie und mehr ein Thema der Philosophie, weshalb sie kein klassisches Thema der Wissenschaftstheorie ist, aber es soll hier gleichwohl darauf eingegangen werden, weil sie (i) oft mit der Frage der Verantwortung für das eigene Handeln verbunden wird, und sie (ii) nach Ansicht einiger Philosophen, aber auch einiger Wissenschaftler ist die Existenz eines freien Willens sogar Voraussetzung für die Wissenschaft (Zeilinger (2010)). Umgangssprachlich scheint die Bedeutung des Ausdrucks 'freier Wille' klar zu sein, – der Wille eines Menschen ist "frei", wenn der Mensch nicht "unfrei" entscheiden muß, d.h. wenn keine Zwänge die Entscheidung zwischen den jeweils möglichen Alternativen bestimmen. Dementsprechend entscheidet ein Mensch frei, wenn er so entscheiden und handeln kann, wie er will; dies ist u.a. der Kern der Definition der Willensfreiheit durch David Hume, s. unten. Diese Definition führt allerdings sofort auf die Frage, ob ein Mensch frei wollen kann, was er will, und was es überhaupt bedeutet, etwas frei zu wollen. Arthur Schopenhauer (1788 – 1860) konstatierte dann auch

"Der Mensch kann zwar tun was er will, aber er kann nicht wollen, was er will."³ Z 1

In einem Interview für die Zeitung "Die Zeit"⁴ kommentierte der Hirnforscher Gerhard Roth Schopenhauers Arbeit zum freien Willen, – sie sei

"wohl die beste [ist], die je über den freien Willen geschrieben wurde. Der Mensch hat zwar einen Willen, aber er kann diesen Willen nicht selbst willentlich beeinflussen. Das ist auch logisch unmöglich: Wenn wir unseren Willen beeinflussen könnten – wodurch würde der Wille, der unseren Willen treibt, beeinflusst? Wieder durch einen Willen, einen dritten, vierten, fünften? Schon seit dem Mittelalter haben kluge Menschen dieses Problem der willentlichen Willenssteuerung erkannt". Z 2

² Heidegger, M. (1951) Im Vortrag "Was heißt Denken?". Gemeint ist damit nach Heidegger "Man kann nicht mit den Methoden der Physik sagen, was die Physik ist. Sondern was die Physik ist, kann ich nur denken." Roth (2004) zitiert den Münsteraner Philosophen Joachim Ritter, der der Ansicht war, dass Naturwissenschaftler die natürliche Welt zwar erforschen können, die Interpretation der Daten aber den Philosophen überlassen müssen.

³In: Die beiden Grundprobleme Ethik, Preisschrift über die Freiheit des Willens (1838), S. 98,

⁴<https://www.zeit.de/campus/2008/02/interview-freier-wille>

Wenn der Wille nicht wollen kann, was er will, muß man wohl folgern, dass er nicht frei ist. Der Wille bildet sich, könnte man sagen, aber diese Formulierung ist mehrdeutig: eine Interpretation ist, dass "der Wille" sich selbst entwickelt, d.h. es wird stillschweigend vorausgesetzt, dass der Wille schon da ist und etwas tut, nämlich sich selbst bilden. So sei das nicht gemeint, wird man nun sagen, aber dann muß man folgern, dass irgendetwas den Willen zustandebringt. Bei "Irgendetwas" muß es sich um einen nicht gewollten Prozess handeln, denn wäre der Prozess gewollt, so wäre man in der Falle des Schopenhauerschen infiniten Regresses. Will man sich auf die Definition von 'frei' als frei von Zwängen beziehen, so ergibt sich eine ähnliche Schwierigkeit, denn Entscheidungen werden im Allgemeinen in Bezug auf irgendwelche Kriterien getroffen: man möchte zum Beispiel einen Nutzen maximieren⁵ oder einen Schaden minimieren, und im Falle nicht unterscheidbarer Alternativen wird zufällig gewählt, wobei erklärt werden muß, was mit 'zufällig' gemeint ist, und ob eine Entscheidung frei sein kann, wenn sie zufällig zustande gekommen ist. Die Freiheit der Entscheidung könnte dann darin bestehen, sich für eine Entscheidung oder allgemein für eine Alternative zu entscheiden, für die man sich nicht selbst entschieden hat, für die sich eine Art mentaler Münze oder mentaler Würfel entschieden hat, also für einen Prozess, dessen Ergebnis nicht voraussagbar ist. Damit ist man allerdings wieder bei der Frage angelangt, wie es zu *dieser* Entscheidung gekommen ist. In der Tat hat der Philosoph Harry Frankfurt den Begriff der "higher-order volition", also des *Willens höherer Ordnung* eingeführt (Frankfurt (1971)), der bei der bei Deliberationen für eine bestimmte Entscheidung eine wesentliche Rolle spielen kann⁶. Zum Begriff des Nutzens muß gesagt werden, dass er sich keineswegs nur auf materielle Größen bezieht, der Ausdruck 'Nutzen' ist nur ein verbales Etikett für eine zu maximierende oder minimierende Größe, er kann sich zum Beispiel auch auf moralische Werte oder auf Frankfurts Willen höherer Ordnung beziehen. Man kann sagen, dass das Bezugssystem gewissermaßen die Entscheidungsbildung leitet, und wenn es die Entscheidung für eine "zufällige" Entscheidung ist. In diesem Sinne kann man argumentieren, der Wille sei nicht frei, bzw. dass es nicht mehr sinnvoll sei, von einer freien Entscheidung zu sprechen, weil der Begriff 'frei' nicht mehr definiert ist, d.h. dass die Frage nach der Freiheit des Willens eher ein Scheinproblem im Sinne Rudolf Carnaps (1928) ist, denn irgendein Bezugssystem wird man immer finden. Interessanter könnte es sein, den Prozess der Abwägung – der Deliberation, wie Philosophen sagen – von Gründen für eine bestimmte Entscheidung zu untersuchen. In empirisch orientierten psychologischen Modellen der Entscheidungsbildung kommen dementsprechend auch keine Variablen vor, die die Freiheit der Entscheidung repräsentieren, sie beschreiben vielmehr den Prozess der Deliberation der Alternativen.

⁵Daniel Bernoulli 1789, von Neumann & Morgenstern (1940), vergl. den Wikipedia-Eintrag "Expected Utility"

⁶Eine neuere Diskussion des Frankfurtschen Begriffs des Willens höherer Ordnung s. a. Yajun(2024)

In der Philosophie dominieren zunächst andere Überlegungen. So war Immanuel Kant (1724 – 1804) sehr beeindruckt von der Physik Isaacs Newtons (1642 – 1726). Kant ging davon aus, dass die im Rahmen der newtonschen Physik hergeleiteten Gesetze als Ausdruck des Kausalitätsprinzips grundsätzlich deterministische Gesetze sind; für ihn stand die Gültigkeit des Kausalitätsprinzips außer Frage. Sollte nun das Kausalitätsprinzip auch auf kognitive Prozesse anwendbar sein, so können Entscheidungen nicht frei sein, weil sie dann ja auf deterministisch verlaufenden Deliberationen beruhen. Menschen können sich aber moralisch entscheiden, d.h. sie werden nicht notwendig nur von "niederen" Bedürfnissen getrieben, so dass Kant folgerte, dass das Treffen von Entscheidungen außerhalb des Gültigkeitsbereichs des Kausalitätsprinzips geschieht, nämlich im Reich der *Noumena*, der Gedanken. Dieses Reich ist eine Art zweite Welt, weshalb auch von der Zwei-Welten-Theorie gesprochen wird. Damit nimmt Kant im Unterschied zu Hume eine nicht-naturalistische Perspektive ein, d.h. eine Perspektive, derzufolge Kognitionen und damit Entscheidungen jenseits der naturgesetzlichen Prozesse stattfinden. Diese Position nimmt auch in der zeitgenössischen Philosophie des Geistes einen breiten Raum ein (s. Abschnitt 1.2.3).

Eine alternative Definition des freien Willens geht auf David Hume (1711 – 1776) zurück. Hume war kein Idealist, und auf die Frage nach der Möglichkeit eines allgemeinen Determinismus geht er nicht ein. Handlungen – und damit Entscheidungen – sind nach Hume frei, wenn sie auf der Basis des Willens und der Wünsche des Akteurs, also der handelnden Person, herbeigeführt werden⁷. Wie der Wille zustande kommt, muß für Hume nicht weiter erklärt werden, 'frei' soll einfach nur bedeuten, dass ein Individuum ohne Einschränkung seines Willens handelt und entscheidet. Der Akteur ist damit auch verantwortlich für seine Entscheidungen, denn sie sind eben nach seinem Willen und seinen Wünschen gefällt worden. Natürlich ist diese Definition von 'verantwortlich' nicht unproblematisch, weil sie wegen Schopenhauers Befund (Zitat 1, Seite 3 und Roths Kommentar 2, Seite 3), in einen infiniten Regress führt, so dass die Definition die Frage aufwirft, ob man für seinen Willen verantwortlich sein kann. Hume ist auf diese Frage eingegangen; Russell (2017) liefert eine ausführliche Diskussion der Diskussionen zu dieser Frage, weshalb eine detaillierte Darstellung hier übergangen werden kann. Jedenfalls kann man sagen, dass eine Entscheidung nicht frei ist, wenn sie nicht nach eigenem Willen und eigenen Wünschen gefällt wird⁸. Damit sind freie Entscheidungen mit dem auch von Hume angenommenen Determinismus der physischen Welt kompatibel, – Hume zählt damit zu den *Kompatibilisten*. Eine zweite Klasse von Anhängern der These des freien Willens sind die *Libertarianer*; sie sind von der Existenz des freien Willens überzeugt und lehnen *deshalb* den Determinismus ab. Wieder andere sind vom Determinismus der physikalischen Welt überzeugt und folgern, dass der Wille nicht

⁷"free actions are those that are caused by the agent's willings and desires" (Russell (2017)

⁸in *A Treatise of Human Nature* (2.3.1–2), bzw. in *Enquiry concerning Human Understanding*

frei sei könne, – dies sind die *Inkompatibilisten*.

Der wesentliche Punkt der humeschen Definition ist, dass die Ursache oder der Grund der Entscheidung der Wille (oder ein Wunsch) des entscheidenden Akteurs, also der entscheidenden Person ist. *Wie* dieser Wille zustande kommt wird in Humes Betrachtungen zur Freiheit und Notwendigkeit nicht weiter diskutiert, und Schopenhauers Bemerkung über das Wollen des Wollens (Zitat) 1) konnte Hume nicht kennen, weil Schopenhauer sie erst nach Humes Tod machte.

Die Frage nach dem freien Willen ist bis heute nicht zur Zufriedenheit aller derer, die sich Gedanken über den freien Willen machen, beantwortet worden. Neurowissenschaftler wie G. Roth und W. Singer sowie Sam Robert Sapolsky (2003) und Sam Harris (2012) argumentieren mit Verve gegen die These der Existenz eines freien Willens. Von zwei renommierten, nobelpreisgeehrten Physikern ist der eine überzeugt, dass ohne den freien Willen gar keine Wissenschaft möglich sei, während der andere die These des freien Willens für baren Unfug hält (s. Abschnitt 1.1.1). Ebenso findet man Neurowissenschaftler, die entgegengesetzter Ansicht sind (s. Abschnitt 1.3).

Für viele Philosophen ist der Gedanke, der Wille könnte *nicht* frei sein, völlig abwegig, und einige von ihnen schließen aus der von ihnen erlebten Evidenz der Existenz des freien Willens auf die Wahrheit der These der Freiheit des Willens. Darüber hinaus folgt für einige Philosophen, dass die Psychologie und sogar die Neurowissenschaft gar keine Wissenschaften sein können, weil die Freiheit des Willens impliziere, dass menschliches Verhalten sich nicht durch auf dem Kausalprinzip beruhende Naturgesetze beschreiben lasse, was u.a. impliziere, dass Verhalten prognostizierbar sein müsse, und Prognostizierbarkeit widerspräche der Freiheit des Willens. Wenn empirisch forschende PsychologInnen davon ausgingen, es gäbe eine wissenschaftliche Psychologie, so erlügen sie einer Illusion.

Bei soviel Uneinigkeit könnte man nun einfach feststellen, dass die Diskussion um den freien Willen wohl eine akademische Streiterei ohne Bedeutung für die Realität sei, – doch so einfach ist es nicht. In der neuzeitlichen Philosophie haben insbesondere Kant und auch Hume die Beziehung zwischen der Verantwortung für das eigene Handeln und der Notwendigkeit moralischer Normen in die Debatte um den freien Willen gebracht, die bis heute fortgesetzt wird: Skeptiker, d.h. Denker, die von der These der Existenz eines freien Willens nicht überzeugt sind, sind der Ansicht, dass die Unfreiheit des Willens im Strafrecht mehr als bisher berücksichtigt werden müsse, wobei sie auf neurobiologische und psychologische Befunde verweisen. Aber auch politische, insbesondere wirtschaftspolitische Theorien basieren oft auf der Annahme, der Wille sei frei, folglich sei das Individuum für sich selbst verantwortlich, etc. Ein Grund von mehreren für diese Einstellung ist, dass wir uns im Allgemeinen als in unseren Entscheidungen frei erleben. Beispiele für diesen Ansatz werden in Abschnitt 1.2.3 vorgestellt. Auf gesellschaftspolitische Fragen wird im Folgenden nicht eingegangen, weil schon ein Versuch, dies zu tun, den hier gegebenen Rahmen völlig sprengen würde.

In folgenden Abschnitt wird auf die Rolle der Physik in der Diskussion um die Determiniertheit versus die Nicht-Determiniertheit, d.h. die Freiheit des Willen eingegangen, insbesondere auf die Frage, ob die Physik notwendig durchgängig kausal strukturiert ist oder nicht. Der häufige Verweis auf die Heisenbergsche Unschärferelation, aus der die für die freien Entscheidungen angeblich wichtige Stochastizität hervorginge spielt für die Anhänger eines totalen Determinismus keine Rolle. Es soll deswegen noch einmal kurz auf die Problematik des schon in Kapitel IV (1) diskutierten Determinismus eingegangen werden.

1.1.1 Physik, Kausalität, und freier Wille

In Wissenschaftstheorie IV (1) – Kausalität und Zufall, Abschnitt 2, ist die Problematik des Determinismus bereits diskutiert worden, trotzdem soll sie hier noch einmal darauf zurückgekommen werden, weil sie für die Hypothese, der Wille sei frei, relevant ist. Es war der Physiker Werner Heisenberg (1923), der zuerst sich selbst und dann eine Mehrheit der Physiker davon überzeugt hatte, mit seiner Unschärferelation die "Kausalität abgeschafft" zu haben (v. Weizsäcker (1999), p. 315). Es mag nützlich sein, das Argument W. Heisenbergs zu rekapitulieren:

"Aber an der scharfen Formulierung des Kausalgesetzes: "Wenn wir die Gegenwart genau kennen, können wir die Zukunft berechnen", ist nicht der Nachsatz, sondern die Voraussetzung falsch, denn wir können die Gegenwart in allen Bestimmungsstücken prinzipiell nicht kennenlernen. Deshalb ist alles Wahrnehmen eine Auswahl aus einer Fülle von Möglichkeiten und eine Beschränkung des zukünftig Möglichen. Da nun der statistische Charakter der Quantentheorie so eng an die Ungenauigkeit aller Wahrnehmung geknüpft ist, könnte man zu der Vermutung verleitet werden, dass sich hinter der wahrgenommenen statistischen Welt noch eine "wirkliche" Welt verberge, in der das Kausalgesetz gilt. Aber solche Spekulationen scheinen uns, das betonen wir ausdrücklich, unfruchtbar und sinnlos. Die Physik soll nur den Zusammenhang der Wahrnehmungen formal beschreiben. Vielmehr kann man den wahren Sachverhalt viel besser so charakterisieren: Weil alle Experimente den Gesetzen der Quantenmechanik und damit der Gleichung (1)⁹ unterworfen sind, so wird durch die Quantenmechanik die Ungültigkeit des Kausalgesetzes definitiv festgestellt." Heisenberg (1927) , p. 175, Z 3

Heisenbergs Argumentation ist nicht nur deswegen bemerkenswert, weil sie eine radikale Aussage über das bis dahin sakrosankte Kausalitätsprinzip ist (von

⁹Es geht in Heisenberg (1927) (p. 174) um die Frage der Ortsbestimmung eines Elektrons, wozu Licht mit der Wellenlänge λ verwendet wird. Es sei p der Impuls (= Masse mal Geschwindigkeit) eines Teilchens (Korpuskel), h sei die Plancksche Konstante; dann gilt $\lambda = h/p$. Der Ort q eines Teilchens ist mit einer Genauigkeit $\Delta q > 0$ bekannt, p mit einer Genauigkeit Δp ; dann gilt $\Delta p \Delta q \geq h$, oder, in Heisenbergs (1927) Notation $\Delta p \Delta q \sim h$. Dies ist die Beziehung (1) auf Seite 175 von Heisenberg (1927)

Aristoteles bis Leibniz: Jedes Ereignis hat seine Ursache), sondern auch weil es eigentlich keine physikalische Begründung für seine Interpretation der experimentellen Daten gibt. Heisenberg will eine metaphysische Aussage über eine wirkliche Welt, in der das Kausalgesetz gilt, vermeiden, weil metaphysische Aussagen sinnlos seien, und postuliert: "Die Physik soll nur den Zusammenhang der Wahrnehmungen formal beschreiben". Das ist ein wissenschaftstheoretisches, der positivistischen Position Ernst Machs und des (frühen) Wiener Kreises entsprechendes Argument. Noch im Jahr 1958 führt er aus:

Einer bestimmten Wirkung eine bestimmte Ursache zuzuordnen hat nur dann einen Sinn, wenn wir Wirkung und Ursache beobachten können, ohne gleichzeitig in den Vorgang störend einzugreifen. Das Kausalgesetz in seiner klassischen Form kann also seinem Wesen nach nur für abgeschlossene Systeme definiert werden. In der Atomphysik ist aber im Allgemeinen mit jeder Beobachtung eine endliche, bis zu einem gewissen Grade unkontrollierbare Störung verknüpft, wie dies in der Physik der prinzipiell kleinsten Einheiten auch von vorn herein zu erwarten war. Da andererseits jede raum-zeitliche Beschreibung eines physikalischen Vorganges die Beobachtung des Vorgangs voraussetzt, folgt, dass die raum-zeitliche Beschreibung von Vorgängen einerseits und das klassische Kausalgesetz andererseits komplementäre, einander ausschließende Charakteristika des physikalischen Geschehens darstellen. Dieser Sachlage entspricht der Sachverhalt, dass zwar ein mathematisches Schema der Quantentheorie existiert, dass dieses Schema aber nicht als einfache Verknüpfung von Dingen in Raum und Zeit gedeutet werden kann. Durch diese Komplementarität der Raum-Zeit-Beschreibung einerseits und der kausalen Verknüpfung andererseits tritt ferner eine eigenartige Unbestimmtheit des Begriffs "Beobachtung" auf, indem es der Willkür anheimgestellt bleibt, welche Gegenstände man zum beobachtenden System rechnen oder als Beobachtungsmittel betrachten soll.

(Heisenberg 1958/2001, p. 48),

Z 4

Einstein stand bekanntlich der probabilistischen Interpretation der Quantenphysik Heisenbergs skeptisch gegenüber, – es müsse versteckte Größen (hidden variables) geben, die dem Kausalgesetz entsprechend das physikalische Geschehen bestimmen, und 1964 publizierte J. S. Bell die dann nach ihm benannte Bellsche Ungleichung, die einen Test der Kopenhagener Deutung der Quantenmechanik ermöglicht; die in den frühen 70er Jahren durchgeführten Experimente (Freedman & Clauser (1972)) lieferten Daten, die als die Widerlegung der einsteinschen Bedenken gelten (s. a. Kapitel IV, Determinismus, wo ausführlicher auf diese Diskussion eingegangen wird). Allerdings sind diese Ergebnisse anscheinend nicht so eindeutig, wie man den Eindruck haben kann. Bei der Interpretation der Daten anhand der Bellschen Ungleichung hat man anscheinend übersehen, dass in die Herleitung der Ungleichung eine Annahme eingeht, die nicht gelten muß (Brans(1988), Larsson (2014)). Sollte eine besonders strenge Theorie des Determinismus gelten, nämlich dass alles Geschehen vom Beginn des Universums an bestimmt ist (dies ist die Annahme des *Superdeterminismus*),

so haben Rauch et al. (2018) geschätzt, dass seit mindestens 7.8 Milliarden Jahren jedes Ereignis vorherbestimmt ist (die Autoren sind keine Vertreter der Annahme des Superdeterminismus, sie zeigen nur die Implikationen dieser Annahme auf). Hossenfelder & Palmer (2020) sehen allerdings in der Theorie des Superdeterminismus eine Möglichkeit, die Probleme der QM zu überwinden, die durch Bells Annahmen zementiert werden: es ist aus Hossenfelder et al.s Sicht die Annahme der statistischen Unabhängigkeit der versteckten Variablen einerseits und der Messeinstellungen andererseits. Die Unabhängigkeit dieser Größen sei eine intuitive Grundannahme der Physik, weshalb der Superdeterminismus üblicherweise schnell verworfen werde. Die Autoren verweisen auf die Binsenwahrheit, dass Intuitionen nicht notwendig wahr sind. Lässt man die Annahme der Unabhängigkeit fallen, so kann man den Superdeterminismus akzeptieren, der aber nach Hossenfelder et al. die Freiheit des Willens ausschließt. Insbesondere Hossenfelder ist eine nachgerade radikal anmutende Vertreterin der These, dass der Wille nicht frei ist. In ihrem Buch *Existential Physics* (2022) argumentiert sie allgemeinverständlich, also ohne auf die Details der Bellschen Ungleichung einzugehen, für die Determiniertheit des menschlichen Willens: "the future is fixed except for occasional quantum events that we cannot influence" (p. 126) Ihr Argument für diese Auffassung ist die eher rhetorische Frage, wie man denn den freien Willen als mit den Naturgesetzen kompatibel ansehen könne, wenn diese doch einen Determinismus implizieren. Eine Antwort könnte sie bei den Philosophen Julian Nida-Rümelin, Jürgen Habermas und insbesondere bei Alva Noë finden, dessen Superdualismus man als Gegenpol zum Superdeterminismus sehen kann, vergl. Abschnitt 1.2.3.4, ob sie aber deren Thesen überzeugend findet ist eine andere Frage.

Der niederländische Physiker Gerard t'Hooft (1946 –) (Nobelpreis 1999 für seine Arbeit auf dem Gebiet der Quantenfeldtheorie) vertritt eine ähnliche Ansicht:¹⁰

"Whether Divine Intervention takes place or not, and whether our actions are controlled by "free will" or not, will never be decidable in practice. This author suggests that, where we succeeded in guessing the reasons for many of Nature's laws, we may well assume that the remaining laws, to be discovered in the near or distant future, will also be found to agree with similar fundamental demands. Thus, the suspicion of the absence of free will can be used to guess how to make the next step in our science."
-Gerard 't Hooft, 1999 Nobel Laureate in Physics, Z 5

t'Hooft hat sich wiederholt gegen die These des freien Willens geäußert (vergl. t'Hooft (2017, 2020), und seine Argumente beruhen – ebenso wie die von Hossenfelder et al. – nicht auf intuitiven Mutmaßungen.

Der österreichische Physiker Anton Zeilinger (1945 –) (Nobelpreis 2022 für seine Arbeit auf dem Gebiet der Quanteninformation und der Verletzung der

¹⁰Die folgenden Zitate von t'Hooft und Zeilinger sind "Reddit PhilosophyofScience, Free Will required for Science or not" entnommen worden.

Bellschen Ungeichung) argumentiert dagegen

”The second important property of the world that we always implicitly assume is the freedom of the individual experimentalist. This is the assumption of free will. It is a free decision what measurement one wants to perform. In the experiment on the entangled pair of photons, Alice and Bob are free to choose the position of the switch that determines which measurement is performed on their respective particles. It was a basic assumption in our discussion that that choice is not determined from the outside. This fundamental assumption is essential to doing science. If this were not true, then, I suggest, it would make no sense at all to ask nature questions in an experiment, since then nature could determine what our questions are, and that could guide our questions such that we arrive at a false picture of nature.” In: Zeilinger (2010). Z 6

Wenn zwei renommierte Physiker zu derart unterschiedlichen Aussagen kommen, kann man vermuten, dass sie nicht denselben Begriff eines ’freien Willens’ haben. Es ist hier nicht möglich, in die Details der Theorien zu gehen, die möglicherweise hinter diesen Aussagen stehen, aber unterschiedliche Definitionen des freien Willens lassen sich schon diskutieren. Bevor auf diese Frage eingegangen wird, sei noch einmal auf grundsätzliche Aspekte des Begriffs des Determinismus eingegangen: Das Konzept scheint sich einfach genug aus dem Kausalitätsprinzip ’Jedes Ereignis hat seine Ursache’ zu ergeben¹¹.

Aber so einfach ist die Sache nicht. Bereits David Hume (1789) hat sich skeptisch über die Begriffe Ursache und Wirkung geäußert, obwohl man in der empirischen Forschung kaum ohne sie auskomme. Aber es handele sich um Konstruktionen des Verstandes, so dass ihnen eine gewisse Beliebigkeit eigen sei. Bertrand Russell hat in seinem Essay ”On causation” den Begriff der Kausalität gewissermaßen aufgelöst, und John Earman liefert in seinem Buch *A Primer on Determinism* (1986) weitere Einblicke in die Komplexität dieses Begriffs; in Kapitel IV¹², insbesondere Abschnitt 2.4 ist bereits darauf eingegangen worden. Wie Earman zeigt ist schon die newtonsche Physik nicht das Paradies für den Determinismus, für das sie normalerweise gehalten werde: der Begriff des Naturgesetzes sei ”crucial, but troublesome”.

van Kampens Betrachtungen: Quantenfluktuationen werden von Philoso-

¹¹Es sollte hier auf Conway & Kochens Arbeit *The Free Will Theorem* (2006) zumindest hingewiesen werden, in der die Autoren beweisen, dass, wenn Experimentatoren eine Entscheidung über eine vorzunehmende Messung fällen, die frei ist in dem Sinne, dass die Entscheidung keine Funktion irgendwelcher den Experimentatoren verfügbaren Informationen über die Partikel ist, dann das Verhalten vermessenen Partikel ebenfalls keine Funktionen irgendwelcher, vor der Messung verfügbarer Informationen sind. Der Beweis beruht auf Axiomen, deren Wahrheit evident sei. Der freie Wille der Experimentatoren wird hier vorausgesetzt, nicht bewiesen. Damit machen sie u.a. eine Aussage über die Beziehung zwischen physischer und mentaler Welt: Sollte für die physische Welt ein strikter Determinismus gelten, so kann der Wille nicht frei sein. Das Theorem zielt, wie schon die Bellsche Ungleichung, auf die Nichtexistenz versteckter Variablen; Kritik wurde zum Beispiel von Goldstein et al (2009) formuliert.

¹²Wissenschaftstheorie IV – Kausalität und Zufall

phen oft angeführt, um sich nicht einem strengen Determinismus ausgesetzt zu sehen. In diesem Zusammenhang ergibt sich die Frage, ob diese Fluktuationen überhaupt auf dem Aktivitätsniveau von Neuronen relevant sind, weshalb noch einmal auf van Kampens Arbeit "Determinism and predictability" (1991), p.273, eingegangen werden soll. van Kampens Analysen beziehen sich auf den Laplaceschen Determinismus¹³, aber sie gelten im Prinzip auch für die Prozesse im atomaren bzw subatomaren Bereich. Es wird ja angenommen, dass die Interaktionen zwischen Teilchen zunächst einmal deterministischen Gesetzen folgen, schließlich ist Schrödingers Gleichung eine deterministische Gleichung. Aber wie weit schlagen Fluktuationen in diesem Bereich durch? Betrachtet man molekularbiologische Vorgänge, so sind sie relativ zu den subatomaren gewissermaßen grobkörnig, aber sie – etwa Stoßprozesse – werden wieder durch deterministische Gleichungen beschrieben, auch wenn diese im Prinzip Approximationen darstellen. Auf makroskopischem Niveau wie in der statistischen Mechanik wiederholt sich das Prinzip. Die tatsächlich beschriebenen Variablen zeigen stochastische Fluktuationen. Die Gleichungen beschreiben die durchschnittliche Dynamik, und wie im Gesetz der Großen Zahlen beschrieben sind die Fluktuationen klein relativ zum Durchschnitt der Fluktuationen. Auf der mesoskopischen Ebene erhält man stochastische Prozesse und die korrespondierenden Differentialgleichungen liefern Wahrscheinlichkeiten, wobei die Verteilungen für Zustände oft gegen Gauss-Verteilungen konvergieren. In dem Maße, in dem analoge Ansätze für (Teil-)Populationen von Neuronen gelten, kommt es zu "rauschinduzierten" Zustandsübergängen, d.h. zu Übergängen, die nicht deterministisch, sondern eben nur probabilistisch beschreibbar sind (Horsthemke & Levever (1984)). Geht man davon aus, dass Kognitionen, insbesondere Entscheidungsprozesse das Resultat von Hirnprozessen sind, so kann man vermuten, dass derartige Zustandsänderungen auch Entscheidungen beeinflussen können. Aber man kann so zustande gekommene Entscheidungen kaum "frei" nennen, weil der Mensch sie eben nicht "frei" getroffen hat, denn sie sind ja durch Fluktuationen, über die er keine Kontrolle hat, beeinflusst werden (Rubinov, Sporns, & Breakspear (2011)), die Rede ist von neuronalen Bursts durch selbst-organisierte Kritikalität (Bak, Tang & Wiesenfeld (1988)). Bei diesen Betrachtungen werden die von der Unschärferelation implizierten zufälligen Ereignisse nicht genannt, allerdings tauchen sie in der im Folgenden referierten Arbeit von Ellis auf.

George F.E. Ellis: From Chaos to free will George F.R. Ellis¹⁴ Ellis (2020) liefert eine Beschreibung der neuronalen Aktivität, die von Anfang an darauf abzielt, die Zufälligkeit in dieser Aktivität zu betonen. Bei Problemstellungen etwa im Rahmen der klassischen Mechanik gehen Physiker von dem mathematischen Sachverhalt aus, dass die Dynamik des Systems durch die Startbedingungen bestimmt ist, – im Prinzip zumindest. Das heißt aber nicht, dass notwendigerweise

¹³van Kampens Theorem: "Der ontologische Determinismus à la Laplace kann anhand von empirischen Beobachtungen weder bewiesen noch widerlegt werden." Beweis: van Kampen (1991), oder Wissenschaftstheorie IV – Kausalität und Zufall, p. 44

¹⁴George R. R. Ellis, (1939 –), Mathematiker, Kosmologe

exakte Vorhersagen möglich sind, wie schon im Zusammenhang mit dem Laplaceschen Dämon erwähnt wurde. Viele Philosophen beziehen sich auf die Nichtvorhersagbarkeit auch deterministischer dynamischer Prozesse als Ausgangspunkt für eine Argumentation für den freien Willen, aber Unvorhersagbarkeit bedeutet ja nicht Freiheit der Entscheidungen. Die Nichtvorhersagbarkeit tritt schon bei relativ einfachen dynamischen Abläufen auf: (i) das Dreikörperproblem: Die Bahnen von drei umeinander kreisenden Himmelskörpern sind nicht mehr mit elementaren Funktionen beschreibbar, das Gleiche gilt für das Doppelpendel, und das Verhalten von drei künstlichen Neuronen in einem künstlichen neuronalen Netz ist unter bestimmten Bedingungen nicht mehr analytisch berechenbar (an der Heiden (1996)), zitiert nach Roth (2014), p. 165. Die Dynamik von Hirnprozessen ist ungleich komplexer als die dieser Systeme, so dass es überhaupt nicht verwundert, dass die Vorgänge im Gehirn nicht im Detail vorhersagbar sind. Andere Philosophen beziehen sich auf die Stochastizität quantenmechanischer Prozesse, – die aber nicht hilfreich ist, denn wenn der Zufall den Willen regiert, kann man ihn kaum als frei ansehen. Ganz abgesehen davon weist Roth (2014) noch einmal darauf hin, dass Prozesse, bei denen quantenmechanische Zufälligkeit eine Rolle spielen könnte (das Ausschütten von Transmittervesikeln an Synapsen, das Öffnen und Schließen von Ionenkanälen in den Zellmembranen von Neuronen, etc) sich ”um Größenordnungen unterhalb der verhaltensrelevanten Ebene” abspielt, – die Komplexität der Hirnvorgänge stellt ein viel größeres Problem dar als der quantenmechanische Indeterminismus (Roth (2014), p. 164). Ellis (2020) jedenfalls ist von der Existenz des freien Willens überzeugt, und so macht er, trotz aller Betrachtungen zur Rolle der Physik, auf Seite 10 einen gedanklichen Sprung und spricht von echtem mentalen Funktionieren:

”Genuine mental functioning and the ability to make decisions in a rational way is a far more persuasive explanation of how books get written”.

Mit den Büchern meint er die Dramen William Shakespears oder die Encyclopedia Britannica: würde der Determinismus alles Geschehen lenken, so wäre die Produktion solcher Bücher schon von Beginn des Universums vorprogrammiert. Das erscheint als absurd, weshalb er ohne weitere Umschweife die Existenz der Fähigkeit zu rationalen Entscheidungen postuliert.

Dagegen wird vorgebracht, dass zumindest aus der Kopenhagener Interpretation der Quantenmechanik folge, dass es einen absoluten Zufall gibt, also einen Zufall, der nicht nur mangelndes Wissen reflektiert. Ellis argumentiert weiter, dass es zwar richtig sei, dass bei deterministischen Differentialgleichungen für gegebene Anfangsbedingungen die Menge der Zustände, die durchlaufen werden determiniert ist, dass aber Störungen von außerhalb des Systems diese Menge verändern. Ellis betrachtet als Beispiel hierfür die Dynamik eines Pendels: einmal ausgelenkt liegt die Bewegung des Pendels fest, – bis es zu einer Störung

kommt, die die Amplitude und die Phase des Pendelbewegung ändert. Ellis kommt dann auf die grundlegenden molekularen Prozesse zu sprechen, z. B: die Aktivität in den Ionenkanälen der Membranen der Neuronen, die sich wegen der Wechselwirkungen zwischen den Neuronen ständig verändert, die die "höheren" Prozesse, die die kognitive Aktivität ausmachen, beeinflussen, und die höheren Prozesse wirken zurück auf die basalen Vorgänge, wo wiederum Heisenbergs Unschärferelation die "unvermeidbare Unschärfe" (unavoidable fuzziness) in den Prozessen erzeugt. Nicht zuletzt die diversen Rückkopplungen in dem hochkomplexen System, das das Gehirn ist, stören einmal begonnene Teilprozesse, die dann eben nicht mehr den Anfangsbedingungen entsprechend weiterlaufen;

"... because of what happens at the microscale. At molecular scales, the processes at work forget initial data due to billions on billions of collisions between molecules every second. Biology thrives on that disorder – a molecular storm ..." (p. 9), Z 7

Ellis verweist dann auf Arbeiten aus der Chaosforschung, in der genauere Vorstellungen, wie sich aus den "molekularen Stürmen" geordnete Strukturen entwickeln können hergeleitet werden. Aus dem Begriff des Determinismus könne diese Erklärung nicht entwickelt werden¹⁵, – und so kommt er zu der Aussage, dass die "außergewöhnliche hierarchische Struktur" der Hirnprozesse diese geordnete kognitive Aktivität und damit auch freie Entscheidungen ermögliche:

"If you seriously believe that fundamental forces leave no space for free will, then it's impossible for us to make choices as moral beings. ... That's a devastating conclusion. We can be grateful it's not true." (p. 10), Z 8

Diese Schlußfolgerung ist – wie schon bei Kant – ein gedanklicher Sprung, von dem man nicht sieht, warum er aus den vorangegangenen Äußerungen folgen soll. Ellis argumentiert, ein unfreier Wille impliziere die Unmöglichkeit moralischen Handelns, die Tatsache, dass Menschen moralisch handeln können, impliziert dann die Existenz eines freien Willen. Die Schlußfolgerung ist nicht zwingend, weil die Prämisse "ein unfreier Wille impliziert die Unmöglichkeit moralischen Handelns" nicht korrekt sein muß, und mit großer Wahrscheinlichkeit auch nicht korrekt ist, denn Ellis übersieht, dass der Wille einer Person unfrei sei kann, weil erlernte Normen ihr bestimmte Entscheidungen mit hoher Wahrscheinlichkeit unmöglich machen. Derartige Normen fallen bei Dretske (1989) unter die Kategorie *Gründe* und werden von vielen Philosophen genannt, um die Natur der Willensfreiheit zu erklären (s. Abschnitt 1.2.3).

Ellis stellt überhaupt nicht in Rechnung, dass die von ihm erwähnten "höheren" kognitiven Prozesse Sozialisierungsprozesse umfassen, die die allermeisten

¹⁵Ellis geht nicht darauf ein, dass mit dem Begriff des Chaos in der Physik i. A. das deterministische Chaos gemeint ist, das in perfekt deterministischen Systemen auftreten kann.

Menschen durchlaufen und die soziales, prosoziales und damit moralisches Verhalten generieren. Diese Entwicklung setzt nicht notwendig einen freien Willen voraus. Man kann argumentieren, dass viele Entwicklungen ein Produkt der Evolution sind: Gruppen von Menschen, in denen sich soziale ("moralische") Normen entwickelt haben, haben eine höhere Überlebenswahrscheinlichkeit als Gruppen, in denen sich keine derartigen Normen entwickelt haben¹⁶. Man könnte mit einem Hauch von Polemik sagen, dass moralisches Verhalten möglich wird, gerade weil der Wille nicht frei ist, denn ein von den Sinnen diktiertes (Kant) asoziales Verhalten kann zwar einen unmittelbaren Vorteil haben: ich schlage meinen Mit-Jäger tot, weil ich dann das erlegte Wild ganz für mich habe und damit kurzfristig meine Überlebenswahrscheinlichkeit erhöhe. Langfristig ist ein solches Verhalten aber nachteilig, weil ich nicht mehr versorgt werde, wenn ich mir auf der Jagd ein Fußgelenk verstaucht habe, also weder jagen noch vor dem Bären fliehen kann und deshalb sterbe. Als Mitglied einer Gruppe von Jägern, die erlegtes Wild miteinander teilen, lebt man länger, weil man wegen eines verstauchten Gelenks oder oder anderer Unbill für eine gewisse Zeit von den anderen Gruppenmitgliedern durchgefüttert wird. Die Einsicht in die Notwendigkeit der Kooperation determiniert das Verhalten, das Postulat, dass die Akzeptanz dieser Norm ein Resultat der Existenz des freien Willens ist, ist gar nicht notwendig, um ein durch Einsicht geprägtes Verhalten zu erklären. Ein Beispiel für die verhaltensbestimmende Kraft von Normen ist das Taboo bei polynesischen Gruppen: hat ein Mitglied einer Gemeinschaft ein Taboo verstoßen, so kann es erkranken, auch wenn noch gar nicht bekannt geworden ist, dass diese Person ein Taboo gebrochen hat, – das Verhalten der Gruppenmitglieder vermittelt den Taboo-Bruch, ohne dass er explizit genannt wird¹⁷. Jedes Mitglied der Gemeinschaft wird sehr darauf achten, keines der verschiedenen Taboos zu brechen (Taboos werden zum Beispiel für eine beschränkte Dauer ausgesprochen, – z.B. das Ernten von Kokosnüssen – um das Überleben der Gemeinschaft auf einer Insel zu sichern). Man benötigt keine Willensfreiheit, um Taboos zu entsprechen, sie wirken nahezu automatisch.

Man kann natürlich argumentieren, dass die Annahme eines freien Willens mit einsichtigem Verhalten kompatibel ist. Der Punkt ist, dass diese Annahme nicht notwendig aus einsichtigem Verhalten folgt und deshalb dem Ockhamschen Rasierer überlassen werden kann. Es folgt dann, dass weder moralische noch unmoralische Entscheidungen etwas über die Freiheit oder Unfreiheit des Willens aussagen. Wichtig ist aber der Befund, dass Menschen sich als frei erleben, gerade auch dann, wenn sie eine moralische Entscheidung treffen. Nur bedeutet das Erleben von Freiheit noch nicht tatsächliche Freiheit.

¹⁶Folgt man den Superdeterministen, so können wir aus der Tatsache, dass sich Normen entwickelt haben, folgern, dass diese Entwicklung vom Big Bang an "eingeplant" war, will sagen, dass die ablaufenden physikalische Prozesse diese Entwicklungen notwendig implizieren.

¹⁷Lit: R.L Stevenson – Life in Western Samoa.
<https://www.uniaktuell.unibe.ch/2022/eating-taboos-the-cultural-construction-of-disgust/index-eng.html>)

1.1.2 Libets Experimente, Bereitschaftspotential

Bis hierher ist der Begriff des freien Willens nur als Theorie diskutiert worden. Man mag dann fragen, ob sich empiriebasierte Aussagen, d.h. Daten, über die Möglichkeit eines freien Willens finden lassen. Experimente, deren Ergebnisse relevant für die Hypothese der Existenz des freien Willens sein könnten, werden in diesem Abschnitt besprochen.

Geht man von der Annahme aus, dass das Bewußtsein aus der Aktivität des Gehirns entsteht, so muß diese Aktivität durch Integration vieler Teilaktivitäten bestehen. Eine erste Stufe besteht in der Integration von Information der unmittelbaren Umgebung, d.h. aus der Verarbeitung visueller, auditiver, taktiler Informationen, die mit Zentren gekoppelt werden, die diese Informationen interpretieren. Dabei handelt es sich um komplexe Prozesse, die u.a. motorische Aktivitäten implizieren können. Diese Dynamik entwickelt sich in der Zeit – keine Dynamik ist instantan im strengen Sinn des Wortes, und was im Bewußtsein erscheint, wird relevanten Teilen dieser Dynamik nachgeschaltet sein. Man kann also vermuten, dass die Dynamik begonnen hat, bevor sie im Bewußtsein repräsentiert wird. Walter et al. (1964) sowie Kornhuber und Deecke (1965) fanden, dass willentlichen Hand- oder Fußbewegungen negative kortikale Potentiale in der Größenordnung von 10 – 15 μV (Mikrovolt, 1 μV = 0.000 001 Volt) vorausgingen; sie nannten sie *Bereitschaftspotentiale* (readiness potentials). Dieses Potential wächst mit der "Aufmerksamkeit und intentionalen Beteiligung der Versuchsperson und nimmt bei Gleichgültigkeit ab".

Der Physiologe Benjamin Libet (1983, 1985) berichtete Experimente, die mögliche neuronale Aktivitäten nachweisen sollten, die einer bewußten Entscheidung bzw. Handlung vorangehen. Sollte es derlei Aktivitäten geben, so sollten sie im EEG aufscheinen. Die Aufgabe eines Experimentators besteht darin, die Versuchsperson (VP) zu einer von ihr selbst gewählten Zeit eine einfache Handlung durchführen zu lassen. Bei Libet mußte die VP einfach die Hand heben. Insgesamt bestand die Aufgabe der VP darin, (i) auf einen Monitor zu blicken, auf den ein Punkt zu sehen war, der sich auf einem Kreis im Uhrzeigersinn bewegte. (ii) zu einem von der VP frei wählbaren Zeitpunkt sollte sie eine Hand heben. In dem Moment, in dem ihr der Wille oder der Drang (the urge) kam, die Hand zu heben, sollte sie einen Knopf drücken, der die Bewegung des Punktes stoppte. Nun sollte sie den Punkt auf die Position bewegen, der dem Zeitpunkt entsprach, an dem der Wille zum Handheben bewußt geworden war. Nach einer kleinen Pause begann der nächste Versuchsdurchgang, der wie der eben beschriebene verlief. Es wurden nun nicht nur die von der VP angezeigten Zeitpunkte ausgewertet, sondern diese Daten wurden zu den evozierten Potentialen eines gleichzeitig erhobenen EEG in Beziehung gesetzt.

Das wesentliche Resultat der Untersuchung war, dass stets ein der bewußten Entscheidung vorausgehendes Bereitschaftspotential gemessen werden konnte.

te. Hatte die Vp den Zeitpunkt geplant, begann das Bereitschaftspotential im Durchschnitt 1050 msec vor der bewußten Entscheidung, wählte sie die Handlung spontan, so setzte es 550 msec vorher ein¹⁸. Zwischen dem Zeitpunkt des Bewußtwerdens und dem Beginn der Muskelaktivität lagen im Durchschnitt 200 msec.

Dieses Ergebnis legt die Vermutung nahe, dass die Entscheidungen vorbewußt gebahnt werden. In der Realität sind die jeweils gestellten Aufgaben natürlich sehr viel komplexer – soll man sich überhaupt ein Auto kaufen und wenn ja welches, soll man das Fach A studieren oder besser das Fach B, etc., und moralische Entscheidungen erfordern oft längere kognitive Aktivitäten. Man wird hier Hypothesen aufstellen können, die im Prinzip eine durch kognitive Aktivität – die selbst eine neuronale Basis hat – auf andere kognitive Aktivitäten einwirkt, die ihrerseits Entscheidungen bewirken, für die dann dasselbe Prinzip gilt: bevor die Entscheidung bewußt wird geht ein jeweiliges Bereitschaftspotential voran. Ein wichtiger Punkt dabei ist, dass eine bereits "getriggerte" Entscheidungsbildung noch vor der Ausführung einer korrespondierenden Handlung gestoppt werden kann, oder, wie Libet sich ausdrückte, es kann ein Veto eingelegt werden. Jedenfalls wird es bei komplexeren Entscheidungsprozessen zu einer Wechselwirkung zwischen vorbewußten und bewußten Prozessen kommen.

Natürlich sind die Libet-Experimente nicht unkritisiert geblieben. Philosophischerseits ist betont worden, dass die Ergebnisse noch nicht zwingend zeigen, dass mentale Ereignisse wie das Füllen von Entscheidungen Ausdruck neuronaler Aktivität sind, und dass insbesondere die kausale Deutung der Bereitschaftspotentiale durch Libet nicht zutreffen müsse; Braun et al. (2021) haben eine Metaanalyse zu Untersuchungen von Libet durchgeführt und fanden, dass einige von Libet et al.s Befunden "fragiler erscheinen als man im Licht der substantiellen, darauf aufbauenden wissenschaftlichen Arbeiten erwarten würde"¹⁹. Überdies handele es sich nicht um echte Entscheidungen zwischen zwei (oder mehr) Alternativen, da es ja nur darum ginge, eine Hand zu heben oder nicht zu heben. Auf die Details der Kritik muß hier nicht eingegangen werden (Libet hat 2004 ein ganzes Buch dazu veröffentlicht). Wie zu erwarten wurden Folgeuntersuchungen mit verbesserten Versuchsanordnungen publiziert, etwa von Haggard & Eimer (1999), die zu einem differenzierteren Bild kommen: so scheinen die von Libet beobachteten Bereitschaftspotentiale in der Tat keine *kausale* Rolle zu spielen, dafür aber die *lateralisierten Bereitschaftspotentiale*.

We investigated the relation between neural events and the perceived time of voluntary actions or the perceived time of initiating those actions using the method of Libet. No differences were found in either movement related potentials or perceived time of motor events between a fixed movement condition, where subjects made voluntary movements of a single finger

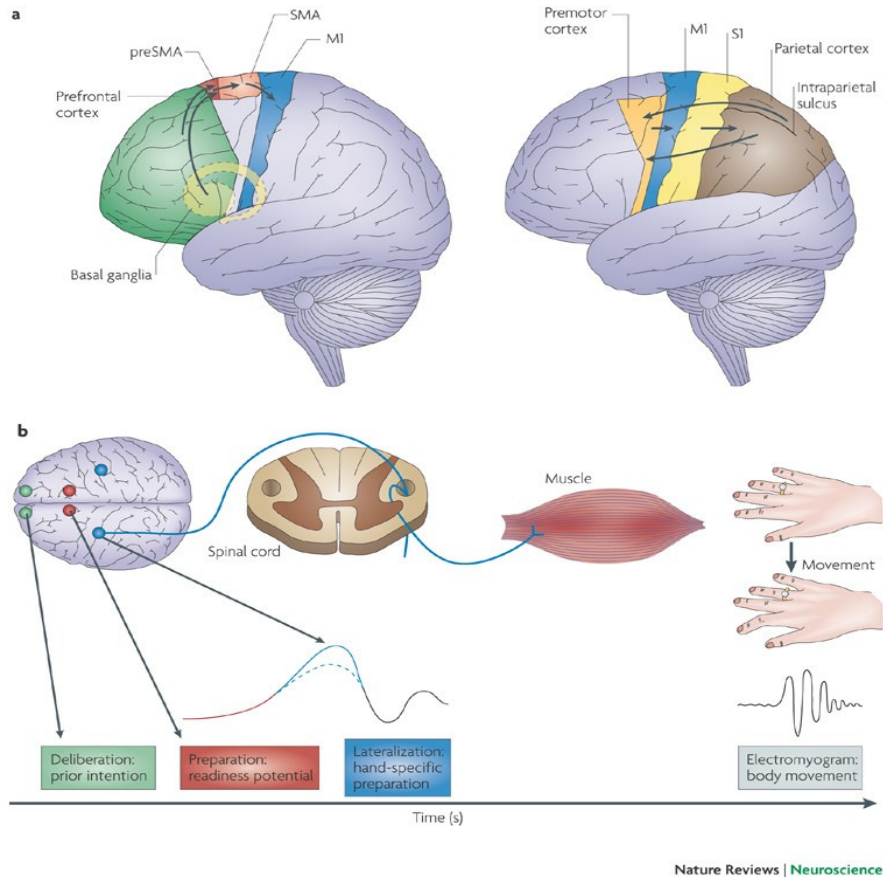
¹⁸1 msec = eine Millisekunde = 1 Tausendstel Sekunde

¹⁹Braun et al.: "We conclude that some of Libet et al.'s findings appear more fragile than anticipated in light of the substantial scientific work that built on them".

in each block, and a free movement condition, in which subjects chose whether to respond with the left or the right index finger on each trial. We next calculated both the readiness potential (RP) and lateralised readiness potential (LRP) for trials with early and late times of awareness. The RP tended to occur later on trials with early awareness of movement initiation than on trials with late awareness, ruling out the RP as a cause of our awareness of movement initiation. However, the LRP occurred significantly earlier on trials with early awareness than on trials with late awareness, suggesting that the processes underlying the LRP may cause our awareness of movement initiation. Z 9

Desmurget et al. (2009) berichten Daten aus neurologischen Untersuchungen, bei denen einzelne Areale des Gehirns durch Elektroden stimuliert wurden, weshalb ein genaueres Bild der neuronalen Aktivitäten entstand, s. a. Haggard (2009). Die Stimulation der rechten inferior parietalen Bereiche erzeugte eine ausgeprägte Intention, die kontralaterale Hand zu heben. Stimulation der linken parietalen Bereiche erzeugte die Absicht, den Mund zu bewegen und zu sprechen. War die Stimulierung hinreichend stark, so waren die Patienten überzeugt, tatsächlich Hand oder Fuß erhoben oder tatsächlich gesprochen zu haben. Die Stimulierung der prämotorischen Bereiche dagegen löste zwar Mund- und Hand/Fußbewegungen aus, aber nun waren die Patienten überzeugt, sich nicht bewegt zu haben (Haggard (2008)). Haggard macht einen wichtigen Punkt: wenn man den freien Willen untersuchen will, kann man nicht einen Stimulus präsentieren und die folgende Response beobachten – Handlungen oder Entscheidungen, die nach freiem Willen geschehen, müssen unabhängig von irgendwelchen Stimuli sein. Die Versuchsanordnung soll also so konzipiert werden, dass nur teilweise vom Experimentator evozierte Entscheidungen vorgegeben werden. Es gibt mehr motorische neuronale Subpopulationen ("circuits"), die an freiwilligen Handlungen beteiligt sein können, die auf den primären Motorkortex M1 konvergieren, vergl. Abb. 1. M1 führt motorische Anweisungen aus, indem er "Anweisungen" an die Muskeln gibt. Soon et al. (2008) lieferten weitere Evidenz für die Hypothese, dass spezifische neuronale Aktivität dem Bewußtwerden einer als frei erlebten Entscheidung vorangeht. Statt EEG- wurden fMRI-Untersuchungen durchgeführt, die die Identifikation verschiedener Bereiche des Gehirns, die an den Entscheidungen beteiligt sind, ermöglichten. Den Vpn wurden Folgen von Buchstaben präsentiert und die Aufgabe war, sobald sie den Drang spürten einen von zwei möglichen Knöpfen mit entweder dem rechten oder dem linken Finger zu drücken. Auf dem Bildschirm erschien eine Anordnung von vier Alternativen, und die Vpn mußte denjenigen Buchstaben angeben, der ihr präsentiert worden war, als sie den Entschluß fassten, den rechten oder linken Finger zu drücken. Danach begann einer neuer Versuchsdurchgang. Die Zeit bis zum ersten Drücken eines Knopfes betrug im Durchschnitt 21.6 Sekunden, – Zeit genug für den Aufbau eines Bereitschaftspotentials. Es zeigte sich, dass beide Knöpfe mit annähernd gleicher Häufigkeit gedrückt wurden. Es konnten spezifische Bereiche des Gehirns identifiziert werden, in denen Aktivierungen beobachtet wurden, die schließlich eine Entscheidung im Bewußt-

Abbildung 1: Lokalisationen für Volitionen (Haggard (2008))



sein manifestierten. Der erste Bereich liegt im frontopolaren Kortex (BA10), 7 ms bevor die Vp die motorische Antwort (Drücken eines Knopfes) produziert wurde. Da die BOLD-Reaktionen²⁰ des fMRI relativ langsam erfolgen vermuten die Autoren, dass die "prädiktive Reaktion" - also der Aufbau eines BP, bereits 10 msec vor der Entscheidung beginnt. Der zweite Bereich liegt im Parietal-kortex, aber schwache Reaktionen wurden auch in anderen Bereichen gefunden. Auf eine detailliertere Beschreibung der Resultate kann hier nicht eingegangen werden, es genügt, zu berichten, dass die Autoren eine mögliche kausale Struktur der Entscheidungsfindung andeuten, derzufolge die Vorbereitung einer Entscheidung im präfrontalen Kortex beginnt und dann in Aktivitäten im supplementärmotorischen Kortex (SMA²¹, ein Teilbereich des motorischen Kortex)

²⁰blood-oxygen-level-dependent (BOLD) signal, detected in fMRI, reflects changes in deoxy-hemoglobin driven by localized changes in brain blood flow and blood oxygenation, which are coupled to underlying neuronal activity by a process termed neurovascular coupling.

²¹SMA: Supplementary Motor Area

übergeht. Der Bewußtwerdung der Entscheidung ging eine nahezu 10 Sekunden dauernde unbewußte Vorbereitung voran, wobei es sich nicht um eine diffuse, sondern um eine spezifische Aktivität handelt, die die spezielle Entscheidung charakterisiert. Soon et al.: "Thus, a network of high-level control areas can begin to shape an upcoming decision long before it enters awareness".

Fried et al. (2011) haben Libet-inspirierte Untersuchungen im Rahmen von neurochirurgischen Eingriffen durchgeführt, indem sie die Aktivität von insgesamt 1019 Neuronen im motorischen Kortex bei selbst-induzierten Fingerbewegungen bestimmt haben. Es wurde also die Aktivität einzelner Neurone direkt durch Elektroden gemessen, während die Vpn (in diesem Fall Patienten) Angaben über ihre bewußten Intentionen gemacht haben. Anhand der Daten konnten Fried et al. Vorhersagen über die Entscheidungen der Personen berechnen 1000 ms bevor der Drang ("urge") bzw. der Wunsch zu handeln bewußt wurde. Die Rekrutierung der Neurone insbesondere im SMA, dem Prä-SMA und dem anterior cingulären Kortex geschah dabei in entweder aktivierender oder inhibierender Weise. Die Daten bestätigen die EEG-Befunde, sind aber einerseits sehr viel genauer und lassen andererseits eine viel detailliertere Analyse als die EEG-Daten zu.

In der Tat haben sich Psychologen schon länger mit der Frage beschäftigt, wie Entscheidungen getroffen werden, ohne dabei speziell "freie" Entscheidungen im Blick zu haben. Die allgemeine Annahme ist, dass stets nach Maßnahme irgendwelcher Kriterien entschieden wird. Für "einfache" Entscheidungen, wie sie im Rahmen psychophysischer Untersuchungen von den Vpn gefällt werden müssen, gibt es Modelle, die auf die stochastische Natur der Entscheidungsfindung fokussieren, etwa Marley & Colonius (1992), oder Vorberg & Ulrich (1987), bei denen Reaktionszeiten, als Wartezeiten bis zu einer manifesten Entscheidung, die abhängige Variable ist, deren Analyse Aufschuß über Details des Entscheidungsprozesses liefert. Ein sehr allgemeines Modell für die Bildung von Entscheidungen wurde von Busemeyer & Townsend (1993) vorgelegt, bei dem der Entscheidungsprozess im Rahmen der Theorie dynamischer Systeme modelliert wird. Das Modell ist zu komplex, um an dieser Stelle auf seine Details eingehen zu können. Auch hier wird auf die Frage nach der Freiheit der Entscheidungen nicht explizit eingegangen. Soll es jedenfalls ein allgemeines Modell sein, muß es auch als frei empfundene Entscheidungen beschreiben können, und das Modell leistet dies, wenn die Kriterien als persönliches Präferenzsystem spezifiziert werden. Bemerkenswert ist nämlich, dass in allen Modellen der hier angeführten Art die auf Kant zurückgehende Charakterisierung freier Entscheidungen, *nicht* Folge irgendwelcher Ursachen zu sein, gar nicht auftaucht. Gemeinsam ist allen Modellen, dass stochastische Effekte explizit berücksichtigt werden, vergl. Braun (2021), denn gerade bei komplexeren Entscheidungen wird ja nicht durch Anwendung eines rigiden Algorithmus entschieden, sondern oft erst nach längeren Erwägungen ("Deliberationen") verschiedener Aspekte der jeweiligen Situation, z.B.: Soll ich das Angebot auf eine interessante, aber

zeitlich begrenzte Stelle im Ausland annehmen oder aber das alternative Angebot einer eher langweiligen, dafür aber zeitlich nicht begrenzten Position im Inland?), – es sei denn, man *definiert* die Kombination von Such- und Assoziationsprozessen als Algorithmus, was aber kein ernst zu nehmendes Argument wäre. Die Entscheidungsfindung gleicht jedenfalls eher einem allgemeinen *random walk*: es werden Aspekte der Alternativen deutlich, die erst im Suchprozess erkannt werden, etc. Unfreie Entscheidungen sind dann solche, bei denen die entscheidende Person den Eindruck hat, eigentlich keine Wahl zu haben, sich also für eine der Alternativen entscheiden zu müssen. Wesentlich ist die Annahme, dass alle Prozesse eine neuronale Basis haben, von deren Aktivität aber nur ein Teil überhaupt bewußt wird. Dabei muß angenommen werden, dass die hier betrachteten Prozesse eine hierarchische Struktur haben, da die Alternativen oder zumindest Aspekte der Alternativen bewußt werden und Gegenstand bewußter Betrachtungen werden, so dass man sagen kann, dass das Gehirn Ergebnisse seiner eigenen Aktivität bewerten muß. Am Ende muß das Gehirn die Struktur seiner eigenen Aktivität analysieren, und man kann der Frage nachgehen, ob diese Selbstreferentialität ein grundsätzliches Hindernis für die Selbsterkenntnis ist oder nicht. Auf Fragen dieser Art soll hier nicht eingegangen werden, weil sie bereits im Rahmen der Thesen von Lucas (1961) und Penrose (1989, 1995) in Teil 1 diskutiert wurden.

1.2 Einige philosophische Ansätze

1.2.1 Immanuel Kant und David Hume

Immanuel Kant (1724 – 1804) Kants Argumentation basiert auf Begriffen, die er insbesondere in seiner *Kritik der reinen Vernunft* geprägt hat. Zunächst geht es um den Begriff der Vernunft, im Unterschied zu dem des Verstandes. Mit Verstand meint Kant die Fähigkeit zu denken und denken heißt bei Kant Begriffe zu bilden, Aussagen zu formulieren und von Aussagen auf andere Aussagen zu schließen. Vernunft dagegen ist die Fähigkeit, zu apriorischer Erkenntnis zu gelangen, – wozu der Verstand notwendig, aber nicht hinreichend ist. Apriorische Aussagen sind nicht-empirische Aussagen, die notwendig wahr sind und ohne Ausnahme gelten (Tetens (2006), p.36). Kant untersucht die Möglichkeit *synthetischer Urteile*, wie zum Beispiel die Aussage "Der Wille ist frei", oder "Die Seele ist unsterblich". Derartige Aussagen sind nicht analytisch, weil sie sich nicht tautologisch sind, sie sind *synthetisch* bzw. *synthetisch priori*. Kant geht von der Frage nach der Möglichkeit von synthetischen Urteilen a priori aus. So nehmen wir Objekte in Raum und Zeit wahr, und Aussagen über Raum und Zeit sind nicht analytisch, weil sie nicht durch Prädikate definiert sind, aus denen die Struktur des Raumes gefolgert werden kann. Bekanntlich nimmt Kant u.a. an, dass die Struktur des Raums euklidisch ist, d.h. er hält die Aussage, die kürzeste Verbindung zwischen zwei Punkten sei die Gerade, für wahr. Für Kant ist diese Aussage nicht empirisch, – sie folgt nicht aus Beobachtungen, aber sie

ist, so die kantsche Formulierung, eine Bedingung für die Möglichkeit, empirische Aussagen etwa über Bewegungen im Raum zu machen. Die Aussage über die Euklidizität des Raums ist synthetisch a priori. Für Betrachtungen über die Möglichkeit solcher Aussagen führt Kant den Begriff *transzendental* ein:

”ich nenne als Erkenntnis *transzendental*, die sich nicht sowohl mit Gegenständen, sondern mit unserer Erkenntnisart von Gegenständen, sofern sie a priori möglich sein soll, überhaupt beschäftigt.” ((B 25) s. Tetens, p. 34). ... Es geht um ”Bedingungen der Möglichkeit von Erfahrung ... dass wir unsere Erkenntnisfähigkeiten ausüben” (Tetens, p. 35).

Transzendente Aussagen spezifizieren die Bedingungen für empirische Erkenntnis. Tetens (p. 35, Fußnote) weist darauf hin, dass ’transzendental’ u.a. für ’notwendig’ steht. Transzendente Aussagen sind metaphysische Aussagen der reinen Vernunft. Nun sind aber die Begrifflichkeiten, auf denen metaphysische Aussagen beruhen, häufig nicht hinreichend scharf definiert, so dass sich Thesen ebenso deren Negation herleiten lassen; dies sind nach Kant die *Antinomien der reinen Vernunft*. Kant schreibt:

Wenn wir unsere Vernunft nicht bloß, zum Gebrauch der Verstandesgrundsätze, auf Gegenstände der Erfahrung anwenden, sondern jede über die Grenze der letzteren auszudehnen wagen, so entspringen *vernünfteln*de Lehrsätze, die in der Erfahrung weder Bestätigung hoffen, noch Widerlegung fürchten dürfen, und deren jeder nicht allein an sich selbst ohne Widerspruch ist, sondern sogar in der Natur der Vernunft Bedingungen seiner Notwendigkeit antrifft, nur dass unglücklicherweise der Gegensatz eben so gültige und notwendige Gründe der Behauptung auf seiner Seite hat. ((B 448/449), s. Tetens, p. 229).

Die dritte Antinomie zeigt Widersprüchlichkeiten der Struktur der Begrifflichkeit des freien Willen auf:

These: Die Kausalität nach Gesetzen der Natur ist nicht die einzige, aus welcher die Erscheinungen der Welt insgesamt abgeleitet werden können. Es ist noch eine Kausalität der Freiheit zur Erklärung derselben anzunehmen notwendig.

Antithese: Es ist keine Freiheit, sondern alles in der Natur geschieht lediglich nach Gesetzen der Natur.

Die Aufgabe des Philosophen sei es nun, diese Antinomie aufzulösen. Dazu geht Kant einerseits davon aus, dass alles, was in der Natur geschieht, strikt dem Kausalitätsprinzip folgt. Andererseits ist er überzeugt, dass dem Menschen eine ”vernünftige Autonomie” zukommt. Um die Entstehung dieser Autonomie zu erklären, muß er zunächst definieren, was er unter Freiheit versteht. Das Kausalprinzip bedeutet Festlegung natürlicher Verläufe und Prozesse, also Unfreiheit, nach der bekannten Schlußregel *modus tollens*²² folgt dann aus der

²² p und q seien Aussagen, $\neg p$ und $\neg q$ seien die Negationen von p und q . $p \rightarrow q$ stehe für ”Wenn p , dann q . Nach dem *modus tollens* folgt dann $\neg q \rightarrow \neg p$.

Freiheit die Nicht-Kausalität. Ist, nach Kant, eine Entscheidung 'frei', so kann sie nicht kausal bewirkt worden sein (die Anführungsstriche beim Wort frei sollen signalisieren, dass es sich um eine Definition des Begriffs *frei* handelt. Kant schreibt:

"dagegen verstehe ich unter Freiheit, im kosmologischen Verstande, das Vermögen, einen Zustand von selbst anzufangen, deren Kausalität also nicht dem Naturgesetze wiederum einer anderen Ursache steht, welche sie der Zeit nach bestimmte." (B 560/61, vergl. Tetens, p. 244).

Um in den folgenden Betrachtungen den Bezug zu dieser Charakterisierung zu erleichtern soll die von Tetens (p. 245) eingeführte Notation übernommen werden

$$\begin{aligned} \cdots \Rightarrow U_{n+2} \Rightarrow U_{n+1} \Rightarrow U_n \Rightarrow \cdots \Rightarrow U_1 \Rightarrow W & \text{ (natürliches Ereignis) } (1) \\ A \Rightarrow U_n \Rightarrow \cdots \Rightarrow U_1 \Rightarrow W & \text{ (freie Entscheidung), } (2) \end{aligned}$$

wobei $\Rightarrow$ hier für "bewirkt" steht. Die drei Punkte links in (1) stehen für eine beliebig lange Folge $U_{n+1}, U_{n+2}, \dots$ von jeweils vorangegangenen Ereignissen in der natürlichen Welt, die sich irgendwo in den Quantenfluktuationen der Welt vor dem Urknall verliert, während A und (2) für ein Ereignis ohne vorangehende Ursache steht und dem Ereignis W in den meisten Fällen zeitlich nahe ist. A ist eine "unverursachte Ursache". Die Ereignisse U_n bis U_1 sind Ereignisse, die die freie Entscheidung Realität werden lassen, wie etwa eine sprachliche Artikulierung, oder eine motorische Handlung. Dieses Vermögen definiert, was Kant die *Vernünftige Autonomie* nennt, Kant spricht auch von *Kausalität aus Freiheit*, –wohl, weil es von A bis W noch eine kausale Folge von Ereignissen gibt, die aber nur der Ausführung der Handlung bzw. Entscheidung dient. Da es in der Welt der Erscheinungen (der natürlichen, physikalischen Welt) nur Ursache-Wirkungs-Ketten der Art (1) gibt folgt, dass es unmöglich ist, Freiheit in der empirischen Welt nachzuweisen. Es ist, so Kant, die Vernunft, die die Idee einer Spontaneität (= Akausalität) schaffe und die ohne vorangehende Verursachung entsteht. Da Freiheit kein empirischer Begriff ist, folgt nach Kant weiter, dass der Ausdruck "Kausalität aus Freiheit" nur für ein "Gedankending" stehe, Freiheit könne also, da sie kein empirisches Faktum sei, nur eine transzendente Idee sein. Freiheit, da nicht Teil der kausal determinierten Welt der Natur, bedeutet Autonomie, insbesondere, so Kant, eine *Vernünftige Autonomie*, womit die Fähigkeit, ein unverursachtes Ereignis zu erzeugen. Die Vernünftige Autonomie impliziert Kant zufolge

- die Möglichkeit, unverursachte Ereignisse durch Nachdenken zu erzeugen. Die Möglichkeit des Nachdenkens ist eine Komponente der Vernunft, die somit ebenfalls nicht dem Kausalprinzip unterliegt.
- Dies bedeutet, dass das Nachdenken nicht nach naturgesetzlichen Regeln abläuft und mithin nicht allgemeinen Kausalgesetzen folge.

- Da Empirie sich immer auf die Welt der Erscheinungen und damit auf kausale Strukturen bezieht, und da die Freiheit als Produkt autonomer Vernunft der empirischen Welt nicht angehört, kann Freiheit nicht empirisch bewiesen werden.

Zwei-Charakter-Theorie Diese besagt, dass jeder Mensch einen doppelten, nämlich einen empirischen und einen intelligiblen Charakter hat. Als Teil der empirischen Welt verursacht unser empirischer Charakter unsere Handlungen, er unterliegt psychologischen Gesetzmäßigkeiten und kann deswegen von der empirischen Psychologie erforscht werden. Der empirische Charakter ist in die kausale Struktur der empirischen Welt eingebunden, weshalb er als nicht frei erscheint. Andererseits ist derselbe Mensch aber auch Teil der *noumenalen Welt*, d.h. des "Noumenon"²³, des Gedachten, das durch die Tätigkeit des Verstandes erkannt werden kann. Tetens (p. 256)) fasst die kantsche Argumentation so zusammen:

"Handeln wir aus vernünftigen Gründen, so sind wir frei und der Naturkausalität nicht unterworfen. Vernunft ist das Vermögen zur Kausalität aus Freiheit."

Hier deutet sich die berühmte Schlußfolgerung Kants an, dass sich unsere Freiheit in unserer Fähigkeit zu moralischem Handeln manifestiert, worauf gleich noch explizit eingegangen werden wird.

Zwei-Welten-Lehre: Die Begriffe, mit denen Kant den Freiheitsbegriff charakterisiert, sind konstituierende Bestandteile den kantschen Zwei-Welten-Lehre, derzufolge es außer der empirischen Welt noch die Welt der *intelligiblen Dinge an sich* gibt. Diese Welt ist das Noumenon, das nach Kant durch die Aktivität des Verstandes erkannt werden kann. Das Noumenon existiert außerhalb der Kausalität, und da Kausalität eine zeitliche Folge der Ereignisse bedeutet, folgt, so Kant, dass das Noumenon außerhalb der Zeit ist. Dazu gehört auch der Prozess der Apperzeption, das ist eine Interaktion von Verstand und Vernunft mit dem sinnlich Wahrgenommenen, die zu einer Gesamtvorstellung von Wahrnehmung und Gedanklichem führt. Die Apperzeption ermöglicht u.a. dem Menschen, sich bis zu einem gewissen Grade selbst erkennen zu können.

Freiheit und Moral: Wenn das Noumenon, die Welt der Gedanken, frei ist, so könnte man auf die Idee kommen, dass diese Welt beliebige Entscheidungen zulässt. Moralische Regeln und Normen ergeben sich aus der Vernunft. Moralisch verhalten wir uns, wenn wir uns nach den Kriterien der Vernunft und nicht etwa nach Kriterien der unmittelbarer Nützlichkeit oder, wie Kant sich ausdrückt, nach "sinnlichem Antrieb" handeln, je mehr "sinnliche Antriebe" in die Entscheidungen eingehen, desto mehr wird die moralische Qualität der Entscheidung reduziert. Um moralisch handeln zu können, müssen wir uns von den "natürlichen Trieben", d.h. den Kausalitäten der Natur befreien können.

²³ von griech. für 'denken'

Frei entscheiden bzw. handeln wir nur dann, wenn dies nach den Kriterien der reinen Vernunft und der damit verbundenen Einsicht in das jeweils Notwendige geschieht. Die Freiheit des Willens wird somit zur notwendigen Bedingung für moralisches Handeln. Handeln wir also vernunftgemäß und damit moralisch, so sind wir frei von den Zwängen des kausalen Naturgeschehens; nur wenn der Mensch über die transzendente Freiheit verfügt kann er die dunklen Triebe aus der Welt der naturhaften Sinnlichkeit überwinden.

Die hier skizzierte Begrifflichkeit gehört zum Kern des von Kant konzipierten *transzendentalen Idealismus*, zu dem Kants These, dass synthetische Urteile a priori möglich sind gehört. Ausdrücke wie Kausalität der Freiheit, Vernunft, Autonomie etc bezeichnen Kompetenzen des Menschen außerhalb der kausalen Welt. Die Freiheit des Willens wird als notwendig für die Fähigkeit zu moralischem Handeln betrachtet, und da der Mensch zu solchem Handeln fähig ist, folgt, dass er über einen freien Willen verfügt.

Kants Überlegungen zur Freiheit des Willens sind aufwändig und sorgfältig, sein transzendentaler Idealismus als philosophischer Rahmen soll eine begrifflich überzeugende und argumentativ dichte Herleitung der These des freien Willens sein. Aber man kann bezweifeln, ob diese Argumentation so zwingend ist wie die Herleitung des Satzes des Pythagoras aus den Axiomen der euklidischen Geometrie. Kant hat sich nicht mit den Details von Denkenprozessen auseinander gesetzt, er hat nur allgemein erklärt, was er zum Beispiel unter 'vernünftigem Denken' versteht. Die Unterscheidung von analytischen, synthetischen, und synthetisch-apriori-Sätzen und damit eines essentiellen Teils seiner Philosophie ist nach sorgfältiger Analyse in Frage gestellt worden (vergl. Wissenschaftstheorie III(2), Abschn. 2.4.2), sein Postulat, dass das Denken nicht kausalen Gesetzen folgt, wird von ihm nicht explizit begründet, und es ist keineswegs klar, warum das Denken *kein* kausaler Prozess sein soll, – zumal der Kausalitätsbegriff weniger klar ist, als die umgangssprachliche Verwendung des Ausdrucks 'kausal' weniger klar ist, als vermutet. Nach Kant ist Freiheit und damit die Freiheit des Willens kein empirischer Begriff, sondern gehört zu den Noumena, also zur immateriellen Zweiten Welt. Woher Kant die Überzeugung nahm, dass es diese Welt gibt, ist letztlich nicht klar, *für ihn* ist das "einfach" so, man könnte ein Evidenzerlebnis vermuten. Man muß fairerweise anmerken, dass zu Kants Zeiten noch sehr wenig über Hirnprozesse bekannt war. Aber seit Kant gibt es viele medizinische und neurobiologische Befunde, die es wahrscheinlicher erscheinen lassen, dass das Denken ein im Gehirn ablaufender Prozess, vermutlich ein Assoziationsprozess sein könnte, wobei detaillierter angegeben werden muß, was unter 'assoziiieren' verstanden werden soll. Aber selbst wenn derartige Kenntnisse zur Verfügung gestanden hätten, hätten sie nicht notwendig Einfluß auf das philosophische Denken gehabt, wie in den folgenden Abschnitten deutlich werden wird, und welchen Einfluß sie auf die kantsche Philosophie gehabt hätten ist schwer oder gar unmöglich zu sagen, – welche Rolle neuronale Aktivität bei der Entstehung von Bewußtsein hat ist bis heute nicht klar, und

experimentalpsychologische, auf mögliche Gesetzmäßigkeiten in Denkprozessen zielende Untersuchungen beginnen erst im 20-ten Jahrhundert, – und liefern oft mehr Informationen über die Fehlerhaftigkeit und weniger über die von den Philosophen der Aufklärung so beschworene Vernunft des Denkens (Johnson-Laird, P. N. (1997)), Tversky, A., & Kahneman, D. (1974)). Philosophie, so scheint es, besteht in der Konzeptualisierung von Phänomenen und der Expliktion von Begriffen, wobei sie aber keine streng durchformalisierte deduktiven Systeme von Aussagen liefern kann, von denen schon Gottfried W. Leibniz (1646 – 1716) geträumt hat; wie Kurt Gödel 1931 gezeigt hat, kann es keine logisch konsistenten Aussagesysteme geben, bei denen jede wahre Aussage streng aus einem endlichen Axiomensystem deduziert werden kann. Das heißt, dass es in jedem philosophischen System Aussagen gibt, die nicht streng logisch herleitbar sind, – was natürlich auch für Aussagesysteme in den empirischen Wissenschaften gilt. Darüber hinaus bleibt das Problem, entscheiden zu müssen, welche Axiome denn die wahren sind: was evident erscheint, muß nicht wahr sein, man denke nur an das Axiom Euklids, dass die kürzeste Verbindung zwischen zwei Punkten eine Gerade sei. Somit hat die Philosophie die Freiheit, auch alten Problemen mit neuen Begriffsbildungen zu Leibe rücken zu können. In der Philosophie des Geistes ist es der Begriff des Grundes, mit dem sich im Unterschied zu dem der Ursache die Freiheit des Willens begründen lassen soll, und die These, dass Entscheidungen gar nicht im Kopf getroffen werden. Darauf wird im folgenden Abschnitt näher eingegangen.

David Hume (1711 - 1776) David Hume geht von einer empirisch näherliegenden Konzeption des Begriffs des freien Willens aus. Handlungen – und damit Entscheidungen – sind nach Hume frei (Russell, P. (2017), wenn sie auf der Basis des Willens und der Wünsche des Akteurs herbeigeführt werden²⁴. Der Akteur ist damit auch verantwortlich für seine Entscheidungen, denn sie sind eben nach seinem Willen und seinen Wünschen gefällt worden. Hume unterstellt damit, dass wir wollen können, was wir wollen, – eine Annahme, zu der sich der deutsche Philosoph Arthur Schopenhauer²⁵ später kritisch äußern sollte.

Eine Entscheidung ist nicht frei, wenn sie nicht nach eigenem Willen und eigenen Wünschen gefällt wird (in *A Treatise of Human Nature* (2.3.1–2), bzw. in *Enquiry concerning Human Understanding*). Damit sind freie Entscheidungen mit dem auch von Hume angenommenen Determinismus der physischen Welt kompatibel. Der wesentliche Punkt der humeschen Definition ist, dass die freie Entscheidung nicht akausal entsteht: die Ursache oder der Grund der Entscheidung ist der Wille (oder ein Wunsch) des entscheidenden Akteurs. Wie dieser Wille zustande kommt wird in Humes Betrachtungen zur Freiheit und Notwendigkeit (liberty and necessity) nicht weiter diskutiert.

²⁴free actions are those that are caused by the agent's willings and desires (Russell (2017)

²⁵1788 – 1860

1.2.2 Martin Heisenbergs Befunde

Der Genetiker und Neurobiologe Martin Heisenberg²⁶ (1983) argumentiert auf der Basis empirischer Befunde einerseits und der Betrachtungen Immanuel Kants zur Frage nach dem freien Willen andererseits, dass der Wille frei sei. Im Folgenden wird Martin Heisenberg als Heisenberg_M notiert, um ihn von seinem Vater Werner Heisenberg (Heisenberg_W) zu unterscheiden).

Was Heisenberg_M (1983) von Kant unterscheidet ist Kants Ansicht, dass die für den freien Willen vorausgesetzte Akausalität notwendig impliziert, dass eine nicht im Bereich des Physischen existierende Vernünftige Autonomie bzw. Kausalität aus Freiheit existiert. Denn er gibt an, eine biologisch generierte *initiale Aktivität* U_0 gefunden zu haben, so dass ohne die Annahme einer dualistischen Basis für die Hypothese eines freien Willen argumentiert werden könne. In Bezug auf die in (1) und (2), Seite 22, eingeführte Notation beschreibt die durch $\Rightarrow \dots \Rightarrow$ repräsentierte Folge von Ereignissen nicht mehr die Entscheidungsfindung, sondern die Ausführung der Entscheidung, z.B. die Aktivierung der Muskeln, die den Arm heben, wenn die Entscheidung darin bestand, *jetzt* den Arm zu heben und nicht eine oder 10 Sekunden später.

Heisenberg_M weist darauf hin, dass Entscheidungen nicht notwendig bewußte Entscheidungen sind. Heisenberg_M ist der Ansicht, dass eine Art von initialer Aktivität bereits aus zufälliger Aktivität resultiert, sofern sie nicht durch äußere Einflüsse generiert wird. Diese zufällige Aktivität könne schon bei einzelligen Lebewesen beobachtet werden, etwa bei der Bewegung des Bakteriums *Escheria coli*. Das Bakterium verfügt über ein *Flagellum*, dass in zwei Richtungen um die Längsachse des Bakteriums rotieren kann. Bei der einen Richtung wird eine Vorwärtsbewegung erzeugt, bei der anderen eine andere Bewegungsrichtung generiert. Die Orientierungen werden zufällig gewählt, so dass ein zufälliger Pfad erzeugt wird, auf dem Futter und die richtige Temperatur gefunden werden kann. Der wichtige Punkt dabei ist, dass diese Aktivität nicht das Resultat eines externen Einflusses ist. Bei der Entwicklung eines Organismus entstünde das motorische System *vor* dem sensorischen System. Dies gelte auch für komplexere Organismen, bei denen Verhalten auf der Basis von Modulen erzeugt wird: z.B. der Greifreflex bei Neugeborenen, die rythmische Bewegung der Beine beim Gehen, das Zuschnappen eines Krokodils. Die Entwicklung derartiger Module sei das Resultat zufälliger und gesetzmäßiger Prozesse im Gehirn, deren physiologische Details bisher nur unzureichend untersucht worden seien. Aber es gebe hinreichend viele Hinweise, dass die Entwicklung von Verhaltensweisen nicht nur das Resultat externer Einflüsse sei. Heisenberg_M verweist auf Untersuchungen an Fruchtfliegen, die ihre Erwartungen über die Konsequenzen ihres Verhaltens auch in Situationen verändern können, die sie in ihrer Evolutionsgeschichte vorher noch nie erlebt haben; Fruchtfliegen können, so Heisenberg_M,

²⁶1940 –

Probleme lösen, vor denen sie noch nie gestanden haben²⁷.

Natürlich stellt sich hier die Frage, was mit einem 'zufälligen Ereignis' gemeint ist. Heisenberg_M verweist auf die Ionenkanäle in Zellmembranen, von deren Zustand es abhängt, wann ein Aktionspotential von einem Neuron ausgelöst wird. Diese Kanäle öffnen oder schließen sich "zufällig", wobei Heisenberg_M u.a. an die Prozesse denkt, die radioaktiven Zerfall ausmachen. Nach Ansicht seines Vaters Werner Heisenberg (Heisenberg_W) folgt aus der von ihm gefundenen Unschärferelation, dass diese Prozesse akausal und in diesem Sinne "zufällig" sind (Akausalität ist keine notwendige Bedingung für Zufälligkeit im Sinne der Wahrscheinlichkeitstheorie), so dass letztlich das Universum nicht deterministisch beschrieben werden kann. Heisenberg_Ms Argumentation ist allerdings ein wenig schnell, denn Heisenberg_Ws Theorie ist zwar Meinungsumfragen unter Physikern zufolge die dominierende Theorie in der Quantenmechanik, aber es gibt Gegenstimmen. Abgesehen davon diskutiert Heisenberg_M nicht die naheliegende Frage, ob stochastische Fluktuationen in den quantenmechanischen Prozessen überhaupt eine Rolle bei eher makroskopischen Prozessen der Entscheidungsfindung spielen. Die empirisch gut belegte Tatsache, dass die Aktivität einzelner Neurone stochastische Komponenten enthält läßt sich auch durch die Zufälligkeit der Stoßprozesse (im Sinne der Nicht-Berechenbarkeit der Trajektorien individueller Moleküle) unter Molekülen erklären, die eine zentrale Rolle bei chemischen Interaktionen spielen und die häufig als Diffusionsprozesse charakterisiert werden können (Turings Reiz-Reaktions-Diffusionsprozesse²⁸). Weiter muß das Skalen-Problem gelöst werden, d.h.h. es muß die Frage nach den Skalen, auf denen die verschiedenen Prozesse laufen, beantwortet werden, vergl. He (2014), Kobaleva et al. (2021)). In Bezug auf den freien Willen muß diskutiert werden, ob er noch frei genannt werden kann, wenn eine Entscheidung von einem zufälligen Ereignis abhängt. Tatsächlich ist die Entscheidung dann ja nicht mehr das Resultat eines Willensaktes eines Individuums, sondern eines Prozesses, der das zufällige Ereignis bewirkt. Die Relevanz dieser Frage ergibt sich aus der Vermutung, dass das Adjektiv 'frei' nicht mehr sinnvoll verwendet werden kann, weil seine Bedeutung im gegebenen Zusammenhang unklar geworden ist. Einerseits werden eher makroskopische Prozesse betrachtet, zB die Evaluation einer Situation in Bezug auf moralische Aspekte, andererseits sollen quantenmechanische Prozesse eine Rolle spielen. Man wird auf die grundlegende Frage zurückgeführt, ob das Naturgeschehen letztlich ein insgesamt deterministischer Prozess ist oder nicht.

Um den Bezug zu den Entscheidungen komplexerer Organismen wie etwa Menschen herzustellen fragt Heisenberg_M nach einer Definition von 'frei' und

²⁷im Original: "For example, my lab has demonstrated that fruit flies in situations they have never encountered, can modify their expectations about the consequences of their actions. They can solve problems that no individual fly in the evolutionary history of the species has solved before. Our experiments show that they actively initiate behaviour."

²⁸Eine sehr klare Darstellung findet man in Murray, J.D.: Mathematical Biology, Springer-Verlag, Berlin, Heidelberg etc 1989

verweist²⁹ dabei auf Kants Ausführungen zur dritten Antinomie: eine Person handelt frei, wenn sie aus eigenem Antrieb tut, was getan werden muß, nur verwendet Heisenberg_M diese spezielle Definition von 'frei' gar nicht weiter, es geht ihm nur um die Akausalität. Tatsächlich ist es denkbar, dass eine Person, die dem kantschen Kriterium entsprechend entscheidet, auch als determiniert betrachtet werden kann: sie hat moralische Vorstellungen entwickelt, die es nicht zulassen, dass sie sich gegen sie entscheidet, und die moralische Entwicklung muß ihrerseits nicht das Resultat einer freien Entscheidung sein, sondern kann ihr durch ein Bestrafungs- und Belohnungssystem aufoktroiert oder zumindest antrainiert worden sein. Dass die introjizierten moralischen Normen es im Allgemeinen erschweren, sich gegen sie zu entscheiden, resultiert aus einem komplexen Wechselspiel emotionaler, kognitiver Prozesse. Letztlich bleibt unklar, warum eine initiale Aktivität Freiheit bedeuten soll, - es ist am Ende eine Frage der Definition des Begriffs 'freie Entscheidung'. Heisenberg_M hat sich entschieden, Freiheit so zu definieren, wie er es eben getan hat, - und ob diese Entscheidung eine freie Entscheidung war oder nicht, bleibt eine offene Frage. Womöglich konnte er gar nicht anders, als sich so zu entscheiden, weil seine begrifflichen Strukturen bereits so angelegt waren, dass es ihm erschien, gar keine andere Wahl zu haben.

1.2.3 Enaktivismus und freier Wille

Theorien zur Existenz oder Nichtexistenz des freien Willens werden im Allgemeinen vor dem Hintergrund allgemeiner philosophischer Annahmen formuliert. Sehr spezielle Annahmen werden in der Klasse der Theorien des *Enaktivismus* (*enacted mind* (Varela et al. (1991)), oder des *erweiterten Geistes* (*extended mind*) (Clark et al. (1998)) gemacht. Kernthesen dieser Philosophien ist das Postulat, dass das Bewußtsein u.a. durch Interaktion mit der Umgebung entsteht. Damit ist gemeint, dass sich Bewußtsein durch eine "transformierende" Interaktion mit der Umwelt ergibt, wobei das Gehirn nicht mehr notwendig der Sitz des Bewußtseins ist. Eine zusammenfassende Übersicht findet man bei Rowlands, Lau, & Deutsch (2020).

Im Folgenden werden zur Illustration dieses Ansatzes kurz einige philosophische, auf enaktivistischen Annahmen beruhende Argumentationen für die Existenz des freien Willens vorgestellt. Wesentlich ist bei ihnen die Zurückweisung der Annahme, dass Entscheidungen im bzw. vom Gehirn getroffen werden. Begründet wird diese Zurückweisung u.a. mit der Aussage, neurobiologische Prozesse seien ihrer Natur nach deterministisch. Entscheidungsprozesse fänden in Wirklichkeit in einem nicht-deterministischen "Raum der Gründe" und nicht in einem kausal strukturierten "Raum der Ursachen" statt und müssten dementsprechend außerhalb des Gehirns ablaufen.

²⁹let's establish what I mean by freedom. One acknowledged definition comes from Immanuel Kant, who resolved that a person acts freely if he does of his own accord what must be done. Thus, my actions are not free if they are determined by something or someone else.

1.2.3.1 Frederik I. Dretske Wie sich dieser Ansatz in philosophischen Arbeiten niederschlägt sei an einer Arbeit des in der Philosophie des Geistes oft zitierten Autors Philosophen Frederik Dretske illustriert³⁰. Er hat aus der Unterscheidung von Gründen und Ursachen (Dretske (1989)) weitgehende Folgerungen gezogen, wobei die Unterscheidung aus einer allmählichen Explikation der Begriffe, insbesondere des Begriffs der Bedeutung besteht. Dretske ist einerseits bekennender Naturalist³¹, der gleichwohl feststellt:

”Meaning certainly isn’t an intrinsic property of meaningful things, something you could discover by looking in the head, by taking the measure of the tracks, or by studying the light under a magnifying glass”, p. 4, Z 10

Damit meint er vermutlich, dass man durch Beobachtung der neuronalen Aktivität eines Gehirns nicht erfährt, welchen bewußten Zustand dieses Gehirn gerade erzeugt, und man deshalb aus der neuronalen Aktivität auch nicht ablesen könne, warum sich ein Mensch in einer bestimmten Weise entscheidet. Er stellt fest,

”Meaning itself is causally inert, powerless to initiate, modify, or influence behavior. The best we can hope for is that it supervenes on something, presumably electrical and chemical events in the nervous system, that packs a real behavioral punch. I find this result quite unacceptable.” (Dretske (1989), p. 4), und ”Even if the things that have meaning are in the head, the meanings themselves aren’t in the head.” (p. 5), Z 11

Die möglicherweise naive Frage ist, wo die Bedeutung denn sein soll, schließlich kann man davon ausgehen, dass das Gehirn bei der Entwicklung von Bedeutungen eine Rolle spielt. Was soll es heißen, wenn Dretske schreibt: ”Die Bedeutung selbst ist kausal inaktiv und kann kein Verhalten initiieren, ändern oder beeinflussen.” Wenn jemand im Haus ruft: ”Es brennt”, ist die Bedeutung unmittelbar klar und den jeweiligen Umständen entsprechende Verhaltensweisen auslösen, – aber derartige Trivialitäten hat Dretske vermutlich nicht gemeint, fragt sich nur, was er denn gemeint hat, – eine Antwort auf diese Frage wird in Dretskes Text nicht geliefert. Es gibt kein Argument für die Aussage, dass Bedeutungen nicht im Kopf sind, es gibt nur Behauptungen, dass es so sei. Es ist wahr, dass wir nicht wissen, wie aus neuronaler Aktivität Bedeutung entsteht, aber die Aussage, die Bedeutung sei außerhalb des Kopfes erklärt noch nicht, *wie* und *wo* es außerhalb des Kopfes entsteht. Es sei an Singers Betrachtung erinnert:

”Die Aufklärung der neuronalen Grundlagen höherer kognitiver Leistungen ist mit epistemischen Problemen behaftet. Eines folgt aus der Zirkularität des Unterfangens, da Explanandum und Explanans eins sind. Das

³⁰Frederick Irwin Dretske (1932 – 2013), US-amerikanischer Philosoph

³¹”I am a materialist who thinks that we sometimes do things because of what we believe ...”, p. 1

Erklärende, unser Gehirn, setzt seine eigenen kognitiven Werkzeuge ein, um sich selbst zu begreifen, und wir wissen nicht, ob dieser Versuch gelingen kann.” (Erster Absatz in Singer (2004), Z 12

Singers Aussage charakterisiert den Kern der Frage, wie Bewußtsein entsteht, impliziert aber nicht die (Hypo-)These, dass es nicht im Gehirn entsteht. Medizinische Befunde und empirische Forschung legen die Gegenthese nahe, nämlich dass das Bewußtsein ein Produkt innerhirnlicher Aktivität ist. Die Behauptung, es sei ein außerhirnliches Ereignis, resultiert dagegen aus reiner Spekulation, deren Charme wohl in der anti-rationalen Negation des Occamschen Prinzips der Sparsamkeit für Theorien liegt: man weiß nicht nur nicht, wie es entsteht, sondern auch nicht, wo es denn entsteht.

1.2.3.2 Jürgen Habermas J. Habermas (1929 –) Habermas ist einer der weltweit meistrezitierten Philosophen und Soziologen der Gegenwart³². Sein Aufsatz ”Freiheit und Determinismus” aus dem Jahr (2004) ist die Ausarbeitung seiner Dankesrede für die Verleihung des Kyoto-Preises 2004, was dazu beigetragen haben mag, dass er, wie Roth (2005) feststellte, zu einer der prominentesten Arbeiten zu diesem Thema wurde. Auf Habermas’ Arbeit wird hier relativ ausführlich eingegangen, weil in ihr ein bestimmter philosophischer Stil besonders deutlich wird, nämlich die Vernachlässigung der begrifflichen Analyse zugunsten der ”gelungenen” Formulierung, die bestimmte Intuitionen akzeptierbarer als deren Gegenintuitionen macht.

Nach Habermas ist der Begriff des freien Willens nicht mit der Idee kompatibel, dass Entscheidungen vom Gehirn getroffen werden, weil das Gehirn ein biologisches Organ ist, dessen Funktionsweise durch Naturgesetze beschreibbar sei, und Naturgesetze sind nach Habermas’ Auffassung deterministische Aussagen. Ein zentraler Punkt der Arbeit ist die These, dass sich mentale Aktivitäten, insbesondere Entscheidungsprozesse, zumindest nicht vollständig auf Hirnprozesse reduzieren lassen. Würden Entscheidungen durch Hirnprozesse bestimmt, so wären sie notwendig unfrei. Habermas ist allerdings davon überzeugt, dass der Wille zumindest partiell frei ist; er erklärt auf Seite 871: Er stellt fest, dass

1. Der Determinismus der biologischen Prozesse und die damit implizierte Unfreiheit des Willens ist mit dem ”alltäglichen Selbstverständnis handelnder Subjekte unvereinbar”, und
2. ”Im Alltag kommen wir nicht umhin, uns gegenseitig und bis auf weiteres die verantwortliche Urheberschaft für unsere Handlungen zuzuschreiben.” Diese Urheberschaft kann durch ”objektivierende Sprache der Neurobiologie nicht ernsthaft infrage gestellt” werden.

Dementsprechend seien Entscheidungsprozesse das Resultat der Abwägung von

³²wikipedia.org/wiki/J%C3%BCrgen-Habermas

Gründen, die anders als Ursachen nicht Elemente eines kausalen Beziehungsgefüges seien.

Die Aussage 1 legt nahe, dass Habermas annimmt, dass das "alltägliche Selbstverständnis handelnder Subjekte" die Annahme eines freien Willen rechtfertigt, und dass die Aussage 2 eine weitere Stütze dieser Annahme ist. Der Rest des Aufsatzes besteht aus Rechtfertigungen und Elaborationen dieser Annahmen. Der Punkt ist, dass diese Aussagen nicht als Hypothesen, sondern als Tatsachenaussagen formuliert werden. Viele Kognitions- und Neurowissenschaftler vermuten aber, dass die erlebte Freiheit eine Täuschung, eine Illusion sei, wobei der angebliche Determinismus der biologischen Prozesse gar keine Rolle spielt. Hinter diesen beiden verschiedenen Sichtweisen verbergen sich Annahmen über die Rolle des Bewußtseins, die Habermas bei seiner Argumentation zu übersehen scheint. Im Zusammenhang mit seiner Diskussion des Epiphänomenalismus und des Reduktionismus weiter unten wird auf hierauf noch zurückgekommen.

Habermas will ohne "metaphysische Hintergrundannahmen" wie etwa die descartessche Zwei-Substanzen-Theorie oder die kantsche Zwei-Welten-Theorie auskommen und die "*richtige Weise* der Naturalisierung des Geistes" (p. 873) finden. Er schreibt:

"Ein Wille bildet sich, wie unmerklich auch immer, im Zuge von Überlegungen. Und weil ein Entschluss erst in der Folge von wie immer auch flüchtigen und unklaren Erwägungen heranreift, erfahren wir uns nur in den mehr oder weniger bewusst vollzogenen Handlungen als frei." (p. 874).
Z 13

Dieser Beschreibung kann man durchaus zustimmen, aber warum daraus folgen soll, dass die Aussage

"Frei nur der überlegte Wille [ist]" (p. 874), und "Der Handelnde ist dann frei, wenn er will, was er als Ergebnis seiner Überlegung für richtig hält. Als Unfreiheit erfahren wir nur einen von außen auferlegten Zwang, der uns nötigt, anders zu handeln, als wir nach eigener Einsicht handeln wollen. Daraus ergibt sich ein Konzept bedingter Freiheit, das beide Momente - einer Freiheit unter Bedingungen - gleichermaßen berücksichtigt." (p. 875),
Z 14

gelten soll, ist unklar. Habermas hält sich mit seiner Definition eng an Hume, weicht aber in der folgenden Elaboration von Hume ab. Der Aussage Z 13 werden auch Skeptiker des freien Willens zustimmen können, denn dass wir unsere Entscheidungen oft als frei erleben ist bekannt. Was Skeptiker der Theorie des freien Willens nicht akzeptieren ist die Folgerung, dass wir deshalb auch frei entscheiden. Genau das scheint Habermas aber anzunehmen: er bezeichnet seinen Begriff von freiem Willen einen "phänomologisch angemessenen" Begriff (p. 872), eben weil Personen *sich als freie Entscheider erleben*. Nach Aussage von Z 14 muß der Handelnde seine Entscheidung aber zusätzlich als richtig empfinden, damit sie eine freie Entscheidung ist. Diese Charakterisierung ist ein wenig

merkwürdig, weil die entscheidende Person nun selbst in Abhängigkeit von ihrer Bewertung ("richtig") darüber entscheidet, ob sie 'frei' oder 'unfrei' ist oder nicht, d.h. sie muß eine Entscheidung über eine Entscheidung fällen. Die Frage ist nun, ob ihre Entscheidung über die Richtigkeit eine freie Entscheidung ist, oder nicht. Habermas' Argumentation impliziert einen infiniten Regress, der seine Theorie in Frage stellt. Schon die griechischen Philosophen fanden, dass die These, die Erde ruhe auf etwas (etwa den Schultern des Atlas) die weitere These impliziert, dass auch Atlas auf "etwas" steht, etc, woraus sie folgerten, dass die Erde auf nichts steht und deswegen frei im Raum schwebt. Analogerweise müsste Habermas also zu dem Schluß kommen, dass der Wille *nicht* frei ist. Die Idee, einen "phänomologisch angemessenen" Begriff von Freiheit begründen zu wollen, erweist sich als Falle, weil die Phänomenologie eher auf die These eines unfreien Willens führt.

Überdies geht Habermas mit keinem Wort auf die häufig vorkommenden Entscheidungen unter Unsicherheit ein, von denen man weiß, dass sie von Abschätzungen – die ja ebenfalls Entscheidungen sind – insbesondere von subjektiven Wahrscheinlichkeiten abhängen, so dass die Bedeutung von "richtig" einigermaßen diffus wird. Allgemein muß man feststellen, dass es bei der Diskussion über den freien Willen eigentlich um die objektive Freiheit des Willens und nicht um das Erlebnis des freien Willens geht, das ja gar nicht in Frage gestellt wird. Habermas rechtfertigt seine Definition mit dem Hinweis, dass Entscheidungen *nicht* auf der Basis eines kausalen Netzwerks von Ursachen getroffen würden, denn "was Handlungen implizit erst zu freien Handlungen macht: der interne Zusammenhang mit Gründen" (p. 873/874). Hier wird sozusagen *en passant*, auf jeden Fall willkürlich, eine Unterscheidung zwischen kausalen Netzwerken einerseits und Gründen andererseits getroffen, ohne das klargestellt wird, warum 'Gründe' keine kausalen Strukturen reflektieren. Habermas Aussage, dass die Entscheidung für die Hypothese des freien Willens ein Resultat philosophischer Reflexion sei, ist leer bis auf Erläuterung seiner Absicht, dass er sich der Frage nach dem freien Willen in *nicht-szientistischer* Weise nähern will (p. 772). Was er unter 'szientistisch' versteht hat er in seinem Buch *Erkenntnis und Interesse* ausgeführt: das Prädikat kennzeichne eine positivistische Auffassung von Wissenschaft, die u.a. impliziere, dass die – seiner Ansicht nach – für die Humanwissenschaften einzig relevante Methode der Hermeneutik als unwissenschaftlich gebrandmarkt würde³³. Zur Norm wissenschaftlicher Arbeit gilt aber das Prinzip der Trennung von Tatsachenaussagen einerseits und Hypothesen andererseits. Dieses Prinzip gilt auch in der Hermeneutik, folgt man den Ausführungen des Philosophen Hans-Georg Gadamer in seinem Klassiker *Wahrheit und Methode*. Obwohl bekennender Hermeneutiker folgt Habermas den Empfehlungen Gadamer aber nicht, seine Aussagen zum freien Willen suggerieren den Status von Tatsachenaussagen, seine Ablehnung empirischer Arbeiten als "Produktionsform von Artefakten" reflektiert seine

³³Eine knappe Darstellung der verschiedenen Positionen findet man u.a. in "Zum Positivismusstreit"; <http://www.uwe-mortensen.de/EmpirismuskritikFSalt.pdf>

schon ideologisch wirkende Gleichsetzung von Wissenschaft und Positivismus, die bereits in seinem Frühwerk *Erkenntnis und Interesse* Ausgangspunkt seiner Argumentationen war.

Habermas gibt keine explizite Definition des Begriffs eines Grundes³⁴, diese ergibt sich implizit im Rahmen von Ausführungen der Art

”Eine Motivation durch Gründe erfordert nicht nur einen rational Stellung nehmenden Akteur, für den Gründe zählen, sondern einen Akteur, der sich durch seine Einsicht bestimmen lässt. ... Im Unterschied zur gewöhnlichen Kausalerklärung erlaubt die rationale Handlungserklärung nicht den Schluss, dass beliebige Personen unter den gleichen Antecedenzen zur selben Entscheidung gelangen würden. Die Angabe von vernünftigen Handlungsmotiven reicht für eine Umformung der Erklärung in eine Prognose nicht aus. Zur verantwortlichen Autorschaft gehört nicht allein die Motivation durch Gründe, sondern das begründete Ergreifen einer Initiative, die sich der Handelnde selbst zuschreibt: Das macht den Akteur erst zum ”Urheber”. (p. 877),

Z 15

Dieser Text führt sofort auf die Frage, warum eine Entscheidung ’frei’ sein soll, wenn sie als ’rational’ klassifiziert werden kann, denn zu Entscheidungen führenden Deliberationen können ja sehr wohl deterministisch ablaufen, egal ob sie als rational gelten oder nicht, zumal die Klassifikation ’rational’ oder ’nicht rational’ vom Blickwinkel des Betrachters abhängen kann. Ebenso kann die Selbstzuschreibung einer Initiative ebensogut in einem deterministischen Rahmen geschehen; dass man sich als ’Urheber’ einer Initiative erlebt bedeutet doch noch lange nicht, dass dieses Erleben einer ’freien’ Entscheidung zum Ergreifen der Initiative entspricht. Vielleicht will Habermas sagen, dass die Deliberation von Gründen dann beendet wird, wenn sich bei der entscheidenden Person der Eindruck, eine Einsicht gewonnen zu haben, einstellt. Das mag so sein, aber letztlich hat man es wieder mit einer Entscheidung über eine Entscheidung zu tun: die entscheidende Person entscheidet – bewußt oder unbewußt – dass sich eine Einsicht gebildet hat. Der Deliberationsprozess stoppt, wenn sich das Erlebnis, eine Einsicht zu haben, einstellt. Das sagt sich einfach, ist aber nicht einfach: was unterscheidet den mentalen Zustand ”ich habe eine Einsicht” vom mentalen Zustand ”ich habe noch keine Einsicht” (also weiter deliberieren)? Woraus *folgt*, dass mentalen Zuständen dieser Art folgt, dass sie *nicht* komplexen neuronalen Zuständen entsprechen? Die Tatsache, dass diese noch nicht explizit spezifiziert werden können, bedeutet noch lange nicht, dass mentalen Zustände keine neuronale Basis haben. Man steht vor der Frage, aus welchem Fundus von gesichertem neurowissenschaftlichen Wissen deduziert werden kann, dass *keine* neuronalen Zustände existieren, die mentale Zustände erzeugen können. Einen solchen Fun-

³⁴Wie Roth (2005) anmerkt geht bei Habermas’ der Gebrauch des Ausdrucks ’Grund’ auf seine Arbeiten zum Positivismusstreit, ”Erkenntnis und Interesse” (1968/1973) und ”Zur Logik der Sozialwissenschaften” (1970), zurück, in denen er den empirischen Ansatz (”Positivismus”) in den Sozialwissenschaften scharf kritisierte.

das gibt es nicht. Habermas' Begriff des Grundes liefert keine Begründung für seine Thesen.

Die Produktion einer Einsicht kann ein assoziativer, determiniert ablaufender Prozess sein. Man muß also fragen, ob eine Entscheidung 'frei' sein kann, wenn sie als Resultat eines deterministischen Prozesses ist. Wenn dieser Prozess nicht determiniert, sondern probabilistisch abläuft, ergibt sich dieselbe Frage, denn die Entscheidung hängt immer noch vom Prozess so, wie er abgelaufen ist ab, die "Freiheit der Entscheidung" hat auch hier keinen Ort. Damit wiederum ergibt sich die Frage, ob die Formulierung vom "begründeten Ergreifen einer Initiative, die sich der Handelnde selbst zuschreibt" eine *freie Wahl* begründen kann. Der auf "schöne", auf jeden Fall beeindruckende Formulierungen fokussierende Philosophiestil Habermas' läßt ihn über derartige Pedanterien hinweggleiten, stellt aber den argumentativen Wert seines Textes in Frage und läßt am Ende die Frage nach der Bedeutung von Gründen offen.

Habermas scheint davon überzeugt zu sein, dass der Mensch über einen freien Willen verfügt, einfach weil ihm der Wille phänomenologisch als frei erscheint, und biologische Prozesse seiner Ansicht nach determiniert sind. Aber selbst wenn biologische Prozesse nicht determiniert, sondern probabilistisch ablaufen, wäre damit noch keine notwendige und erst recht keine hinreichende Bedingung für die Freiheit des Willens gegeben, denn aus neurobiologischer Sicht werden mentale Zustände durch biologische Prozesse erzeugt; die Annahme, mentale Zustände hätten eine immaterielle Basis, liefert nachgerade per definitionem keine Einsicht in ihre Entstehung. Die Annahme impliziert nur die Forderung, dass man sich auf eine phänomenologische Beschreibung von mentalen Zuständen und ihrer Dynamik zu beschränken hat; Versuche, die Immaterialität mentaler Prozesse durch Konstrukte wie die Multiple Realisierbarkeit mentaler Prozesse zu begründen, sind wenig überzeugend geblieben³⁵. Wie auch immer, es erscheint Habermas, folgern zu müssen, dass Entscheidungsprozesse zumindest nicht vollständig auf Hirnprozesse zurückgeführt ("reduziert") werden können. Um die Freiheit des Willens postulieren zu können muß Habermas, um den überall lauernden Determinismen entrinnen zu können, eine Form von Dualismus annehmen. Habermas' Begriff der bedingten Freiheit läßt zu, dass der Deliberationsprozess Bedingungen unterliegen kann, die die Anzahl der möglichen Alternativen bei einzelnen Sub-Entscheidungen einschränken. Diese Einschränkungen können sich aus psychologischen Randbedingungen ergeben: so kann der Aufmerksamkeitsfokus weiter oder enger sein, d.h. er kann mehr oder weniger Aspekte ("Gründe") im Entscheidungsprozess berücksichtigen, etc. Deshalb möchte Habermas eine spezielle Art von Dualismus formulieren, nämlich den *Perspektivendualismus*, den er gelegentlich auch *epistemischen Dualismus* nennt. Eine der beiden Perspektiven bezieht sich auf die Rolle (neuro)biologischer Ursachen, die andere auf die nicht-biologischen Gründe. Zu je-

³⁵Vergl. das Kapitel "Philosophie des Geistes und Wissenschaftstheorie (1) – Funktionalismus und die Reduzierbarkeit mentaler Phänomene

der Perspektive gehöre ein spezifisches *Sprachspiel*³⁶. Als Perspektivendualismus bezeichnet er insbesondere einen Dualismus von naturwissenschaftlicher Sichtweise und einer Sichtweise, die sich "den heute bekannten Naturwissenschaften entzieht" (p. 880). Er schreibt:

Aus der anthropozentrischen Sicht einer Lebensform von problemlösenden Sprach- und Kooperationsgemeinschaften vergesellschafteter Individuen bleiben die beiden Vokabulare und Erklärungsperspektiven, die "wir" der Welt auferlegen, für uns "nicht-hintergebar". Das erklärt die Stabilität unseres Freiheitsbewusstseins gegenüber dem naturwissenschaftlichen Determinismus. Andererseits können wir den organisch verwurzelten Geist nur so lange als eine Entität in der Welt verstehen, wie wir den beiden komplementären Wissensformen keine Geltung apriori zuschreiben. Der epistemische Dualismus darf nicht vom transzendentalen Himmel gefallen sein. (p. 881),

Z 16

Aus der genannten Nicht-Hintergebarkeit mag sich die Stabilität unseres Freiheitsbewusstseins ergeben, gleichwohl wird der Sinn dieses Textes um so unklarer, je genauer man ihn liest. Was soll es bedeuten, dass es sich um eine Stabilität "gegenüber dem naturwissenschaftlichen Determinismus" handelt? Vielleicht meint Habermas einfach, dass das Freiheitsbewußtsein mit der Annahme eines naturwissenschaftlichen Determinismus inkompatibel ist. Dies könnte bedeuten, dass das Freiheitsbewußtsein mit einer naturwissenschaftlichen Beschreibung psychischer Prozesse, die keinen allgemeinen Determinismus annimmt, kompatibel ist. Die in den vorangegangenen Betrachtungen zum Determinismus erlauben ja die Möglichkeit solcher Betrachtungen. Aber, wie in den vorangegangenen Abschnitten ebenfalls schon argumentiert worden ist, bedeuten probabilistische neurobiologische Prozesse noch keine Freiheit des Willens, sondern nur eine gewisse Zufälligkeit der Entscheidungen, und Zufälligkeit bedeutet, dass nun General Zufall und nicht die Person entscheidet. Will man aber argumentieren, dass die Zufälligkeiten der Neurobiologie nicht auf die Ebene des Bewußtseins durchschlagen, so muß man erklären, warum sie das nicht tun. Das ist eine Aufgabe, die Habermas anscheinend gar nicht in den Sinn kommt: seine Formulierungen haben anscheinend den Effekt einer Selbstmesmerisierung, die in seine Annahme von Dualismen mündet. Er hat Recht, wenn er schreibt, dass der epistemische Dualismus nicht vom transzendentalen Himmel fallen darf. Seine vorangehenden Formulierungen liefern aber kein Argument, demzufolge der epistemische Dualismus nicht vom Himmel fällt. Dass wir uns als "frei" erleben, ist ja bekannt, kann aber bekanntlich als Illusion gedeutet werden. Dass "ich" eine Entscheidung treffe heißt doch nur, dass die Prozesse, die das "Ich" erzeugen, Entscheidungen als zum Ich gehörig interpretieren, weshalb aber der Entscheidungsprozess noch lange kein "freier" Prozess sein muß.

³⁶Der Begriff des Sprachspiels wurde von Ludwig Wittgenstein in seinen *Philosophischen Untersuchungen* eingeführt. Ein Sprachspiel ist eine sprachliche Äußerung in einem bestimmten Kontext, der u.a. die jeweilige Bedeutung von Ausdrücken festlegt.

Als Alternative zum Reduktionismus möchte Habermas ein "Bild der Interaktion zwischen einem Gehirn, das den Geist determiniert, und einem Geist, der das Gehirn programmiert"³⁷ entwerfen, das ein Ergebnis philosophischer Reflexion und nicht selbst naturwissenschaftliche Erkenntnis" sei (p. 872). Das Gehirn ist nicht der Geist, dekretiert Habermas, aber was der Geist ist, verrät er nicht. Der hier von Habermas gemeinte Geist ist vermutlich nicht der von Hegel konzipierte objektive Geist, der ja eine Art Zeitgeist und deshalb nicht in einem einzelnen Hirn verortet wird. Auch die Frage, warum philosophische Reflexionen grundsätzlich ertragreicher als naturwissenschaftliche Forschungen sein sollen bleibt offen.

Ausgangspunkt der habermasschen Kritik am Reduktionismus ist der von ihm unterstellte Determinismus von Naturgesetzen. Leider haben sowohl Singer wie Roth diese Prozesse ebenfalls als deterministisch charakterisiert, was von Habermas wohl als Bestätigung seiner Einschätzung der Naturwissenschaften gesehen wurde, obwohl das Prädikat 'deterministisch' für die Argumentationen von Singer und Roth gar nicht wichtig ist und von ihnen – so scheint es jedenfalls – eher metaphorisch verwendet wird.

Zur Illustration seiner empiriekritischen Ansichten verweist Habermas auf die empirischen Befunde der Libet-Experimente, die zur Begründung der Unfreiheit des Willens herangezogen werden. Auf Habermas' Kritik der Libet-Experimente muß hier nicht im Detail eingegangen werden, es genügt, festzustellen, dass nach Habermas derlei Experimente irrelevant sind, weil sie seiner Ansicht nach nur Artefakte erzeugen können; freie Handlungen entstehen eben durch einen "internen Zusammenhang mit Gründen" (p. 873-874) und nicht, weil irgendwelche Bereitschaftspotentiale das Verhalten bestimmen. Habermas argumentiert weiter mit der Rolle der verschiedenen Sprachspiele: Wenn ein Beobachter des "Diskursgeschehens" (wohl: des Entscheidungsverhaltens) beschreiben soll, so muß er eine mentalistische Sprache verwenden, "das heißt in einer Sprache, die Prädikate wie 'meinen' und 'überzeugen', 'bejahen' und 'verneinen' enthält" (p. 875). Er schreibt: "In einer empiristischen Sprache müsste er, aus grammatischen Gründen, alle Bezüge auf die propositionalen Einstellungen von Subjekten, die etwas für wahr oder falsch halten, eliminieren". Dies hieße aber, das mentalistische Sprachspiel auf das biologische zu reduzieren. Das aber könne nicht funktionieren, denn die beiden Sprachspiele können seiner Ansicht nach nicht aufeinander reduziert werden: "Gedanken, die wir im mentalistischen Vokabular ausdrücken können, lassen sich nicht ohne semantischen

³⁷Das ist wieder schön formuliert, doch was will Habermas sagen? Wenn das Gehirn den Geist determiniert und der Geist das Gehirn programmiert, so liegt die Folgerung nahe, dass das Gehirn sich in deterministischer Weise selbst programmiert, – was auch immer das bedeuten soll und was er vermutlich gerade nicht gemeint hat. Jedenfalls propagiert er implizit einen Determinismus, von dem er auf der vorgehenden Seite sagt, dass er "mit dem alltäglichen Selbstverständnis handelnder Subjekte unvereinbar" sei, und im ersten Absatz kritisiert er bereits "die deterministische Auffassung, wonach eine kausal geschlossene Welt für die Freiheit der Wahl zwischen alternativen Handlungen keinen Platz läßt".

Rest in ein empiristisches, auf Dinge und Ereignisse zugeschnittenes Vokabular übersetzen ... weil in die Grammatik der beiden Sprachspiele unvereinbare Ontologien eingebaut sind" (p. 881)³⁸. Das empiristische Sprachspiel erfasst nach Habermas nicht die Rolle der "Vergesellschaftung" des Individuums. Aber mit diesen Argumentationen begibt sich Habermas auf das Glatteis verschiedener Theorien zum Reduktionismus, wie sie im des Kapitel *Philosophie des Geistes* (1), Abschnitt 2.4.3, diskutiert wurden.

Reduktionismus Korrespondierend zu seinen Argumenten zum freien Willen ist Habermas' Kritik des Reduktionismus recht einfach: die Neurobiologie suche nach Gesetzen, und naturwissenschaftliche Gesetze seien deterministisch. Habermas unterstellt die Annahme eines Determinismus für derlei Gesetze ungeachtet der Fülle von kritischen Evaluationen dieser Theorie (s. unter anderem van Kampens Theorem, Abschnitt 1.1.1)³⁹. Habermas' mit vielen Philosophen geteilte Unterscheidung zwischen Gründen und Ursachen setzt voraus, was er eigentlich zeigen müsste, nämlich die grundsätzliche Inkompatibilität von angeblich deterministischen neurobiologischen und nicht-deterministischen mentalen Prozessen. Roth (2005) diskutiert diesen behaupteten Unterschied mit explizitem Bezug auf neuroanatomische und neurophysiologische Befunde. In diesem Zusammenhang geht Roth (2005) auch auf die Behauptung Habermas ein, der Reduktionismus impliziere, Bewußtsein und bewußtes Handeln (freie Entscheidungen sind bei Habermas stets bewußte Entscheidungen) seien Epiphänomene. Das sind Phänomene, die kausal erzeugt werden, selbst aber nicht kausal wirken können. Die diesem Argument unterliegende Vorstellung ist, dass neurobiologische und mentale Prozesse grundsätzlich separate Prozesse sind. Es ist aber nicht klar, warum das so sein soll: geht man von einer systemtheoretischen Vorstellung aus, was immer möglich ist – das Gehirn ist ein Paradebeispiel für ein dynamisches System⁴⁰ –, so wird man zwar sagen können, dass uns bestimmte Prozesse im Gehirn, nämlich diejenigen, die tatsächlich Bewußtsein erzeugen, noch nicht bekannt sind, dass man aber davon ausgehen kann, dass sie von bestimmten neuronalen Aktivitäten "angetrieben" werden, was durch neurophysiologische Untersuchungen, die mit simultan ablaufenden Verhaltensbeobachtungen gekoppelt sind, nahegelegt wird. Man hat es dann insgesamt mit *einem* System zu tun, so dass bewußtes Erleben und Handeln durch bestimmte, dazu korrespondierenden Systemzuständen definiert sind. Wie Roth (2005), p.

³⁸Habermas führt hier eine weitere, auf G. H. von Wright, *Explanation and Understanding*, London 1971, Part II zurückgehende Begründung an, auf die hier nicht weiter eingegangen werden muß, weil der "semantische Rest" letztlich mit der Rolle von Gründen bei der Entscheidungsfindung erklärt wird.

³⁹Tatsächlich erwähnt er in einer Fußnote – 4 auf Seite 871 – probabilistische Gesetze: "Die deterministische These ist unabhängig davon, ob wir Naturgesetze probabilistisch interpretieren. Denn Willkür lässt sich nicht auf Zufall zurückführen." Diese "Begründung" ist natürlich anfechtbar, aber typisch für die apodiktische Wisch-und-weg-Argumentation Habermas'.

⁴⁰Gemeint sind hier Systeme, die sich durch Systeme von Differentialgleichungen beschreiben lassen, wobei gegenwärtig kaum jemand auf die Idee kommen wird, ein derartiges System von Gleichungen für das gesamte Gehirn aufstellen zu wollen.

702, anmerkt, läßt dieses Modell auch die Eigengesetzlichkeit des Mentalen zu, ohne dass damit ein Epiphänomenalismus impliziert würde, da Zustände des Systems auf andere Zustände rückwirken können. Diese Formulierung ist gewählt worden, weil der Epiphänomenalismus die Möglichkeit einer Rückwirkung vom Mentalen auf das Neuronale negiert, sie ist allerdings auch irreführend.

Eine Illustration ist das sehr einfache, durch zwei gekoppelte Differentialgleichungen definiertes System, dass in Wissenschaftstheorie IV, Kausalität und Zufall (1), Abschnitt 1.1 ausführlicher diskutiert wird, s. a. den letzten Abschnitt "Ausblick" in *Philosophie des Geistes und Wissenschaftstheorie (1)*. Für derartige Systeme verliert die Rede von der Kausalität, wie "die Werte von x bewirken die von y " oder umgekehrt, ihre Bedeutung: Kausalität bedeutet ja, dass die Wirkung einer Variablen, etwa x , der Veränderung der anderen Variablen (y) vorangeht. So kann man eine Rückkopplungsschleife betrachten: sie bewirkt, dass bestimmte Aktivierungen nicht beliebig wachsen, sondern gedämpft werden. Andererseits bewirkt die Aktivierung auch die Aktivierung der Rückkopplung, – ohne Aktivierung wird auch keine dämpfende Rückkopplung erzeugt. Tatsächlich beginnt mit der Aktivierung auch der Aufbau einer die Aktivierung dämpfenden Aktivität, was den Begriff einer zeitlichen Folge von Ursache und Wirkung unscharf werden läßt. Tatsächlich finden den Gleichungen zufolge die Wechselwirkungen zwischen den Variablen simultan⁴¹ statt. Für neurobiologische Systeme wird man u. U. noch gewisse Laufzeiten der Signale berücksichtigen müssen, so dass vom zeitlichen Vorangehen und damit des kausalen Effekts einer Variablen vielleicht noch geredet werden kann, was aber für die These des Epiphänomenalismus unerheblich ist.

Ein weniger theoretisches und mehr an Daten orientiertes Argument ergibt sich aus der Frage, ob das Bewußtsein Handlungen, und damit Entscheidungen, *bewirkt*. Habermas schreibt "Der Handelnde ist dann frei, wenn er will, was er als Ergebnis seiner Überlegung für richtig hält." (p. 874) An Schopenhauers kritische Überlegung zur Rolle des Willens scheint er bei der Formulierung dieses Satzes nicht gedacht zu haben. Nach Habermas ist freies Handeln demnach bewußtes Handeln. Aber das Erleben eines *freien* Handelns kann das Resultat einer Täuschung sein, denn es ist möglich, dass der Gedanke an eine Handlung und die Handlung selbst parallel generiert werden und so der Eindruck entsteht, dass die Handlung bewußt geplant oder initiiert wurde. Dieser Eindruck könnte ein Taschenspielertrick des Bewußtseins sein (Wegner (2003)); Wegner berichtet u.a. eine Reihe von Experimenten sowohl vom Libet-Typ als auch im Rahmen von klinischen Untersuchungen, die für diese These sprechen. Derlei Daten werden von Philosophen, insbesondere von Habermas, Nida-Rümelin (s. unten) und anderen gar nicht in Betracht gezogen.

Objektiver Geist und Handlungsfreiheit Habermas will den Begriff des Grundes klären und führt dazu die auf Hegel zurückgehenden Begriffe des sub-

⁴¹Dynamische Modelle für die Interaktion auch von neuronalen Populationen werden u.a. in Murray (1989) vorgestellt.

jektiven und des objektiven Geistes ein. Zu dieser Unterscheidung gelangt Habermas durch eine Betrachtung der Kommunikation zwischen Individuen. Damit diese stattfinden kann, müssen sich die Individuen gegenseitig verstehen, was nach Habermas nur möglich sei, wenn die Individuen über ein "intersubjektiv geteiltes kulturelles Hintergrundwissen" verfügen:

"Ohne Intersubjektivität des Verstehens keine Objektivität des Wissens. Ohne die reorganisierende "Anbindung" des subjektiven Geistes und seines natürlichen Substrats, des Gehirns, an einen objektiven Geist, das heißt an symbolisch gespeichertes kollektives Wissen, fehlen propositionale Einstellungen zu einer auf Distanz gebrachten Welt", . . . , "Erst die sozialisierten Gehirne, die an ein kulturelles Milieu Anschluss finden, werden zu Trägern jener eminent beschleunigten, kumulativen Lernprozesse, die sich vom genetischen Mechanismus der natürlichen abgekoppelt haben" (p. 885), Z 17

Also haben die Gehirne doch Anteil an Entscheidungsprozessen, aber eben über den objektiven Geist. Die Verwendung des Ausdrucks 'objektiver Geist' soll wohl die Idee des freien Willens stützen, hat aber kaum einen argumentativen Effekt: natürlich kennen wir alle soziale Regeln, die unser Zusammenleben bis zu einem gewissen Grad regulieren, wir haben Kenntnisse aus der Wissenschaft, aus der Belletristik, etc, all diese Kenntnisse haben wir uns irgendwann und irgendwie angeeignet. Klar ist auch, dass all dies zumindest in einen Teil unserer Entscheidungen eingeht. Diese Feststellung ist im Kern banal, aber Habermas möchte diesem Sachverhalt in wundersam verschlüsselter Weise philosophischen Adel verleihen:

"Der objektive Geist ist die Dimension der Handlungsfreiheit. Im performativ mitlaufenden Freiheitsbewusstsein spiegelt sich die bewusste Teilhabe an jenem symbolisch strukturierten "Raum der Gründe", worin sich sprachlich sozialisierte Geister gemeinsam bewegen. In dieser Dimension vollzieht sich die rationale Motivation von Überzeugungen und Handlungen nach logischen, sprachlichen und pragmatischen Regeln, die sich nicht auf Naturgesetze reduzieren lassen.", p. 886, Z 18

Der erste Satz wird nicht notwendig klarer, wenn man ihn mehrfach liest, insbesondere weil er nicht sagt, was genau er mit "Dimension der Handlungsfreiheit" meint, – der Leser hat die Freiheit der Interpretation. Wesentlich ist die Aussage, dass wichtige kognitive Funktionen und Aktionen sich nach Habermas' Auffassung nicht auf Naturgesetze reduzieren lassen. Er mag damit recht haben, aber er gibt keine Begründung für diese Aussage⁴². Was deutlich an den

⁴²Insgesamt zeigt Habermas eine Neigung zur sprachakrobatischen Mesmerisierung seiner Leser und wohl auch sich selbst, mit der er die Diffusität oder auch Trivialität seiner Argumente camoufliert. Seine Argumente sind ja i. A. gar keine Argumente, sondern nur Beschreibungen bekannter Sachverhalte, – als Philosophie deklarierte Produkte aus seiner Werkstatt für barocke Sprachgirlanden. Der sehr klar schreibende Philosoph Karl R. Popper hat in einer Reaktion auf die Texte von T. W. Adorno und J. Habermas, die im Rahmen des Positivismusstreits produziert wurden, geschrieben: "Ich kann solche Leute wie Adorno und Habermas

Sätzen in Zitat 18 ist, hat er schon an anderer Stelle seines Textes gesagt, dass nämlich der Wille frei ist, weil Entscheidungen auf der Evaluierung von Gründen beruhen, und Gründe sich auch aus öffentlich Gesagtes und Geschriebenes etc beziehen. Warum dieser objektive Geist nun *die* Dimension der Handlungsfreiheit ist, wird nicht erklärt, – möge sich der Leser doch selbst einen Reim darauf machen, schließlich handelt es sich um eine schöne Formulierung. Weiter ergibt sich die Frage, ob jeder Mensch ein "performativ mitlaufendes Freiheitsbewußtsein" hat, wenn er sich zwischen verschiedenen Alternativen entscheiden muß. Vielleicht meint Habermas, dass man sich stets (= "mitlaufend") als frei empfindet, wenn man Gründe deliberierend (= "performativ", oder vielleicht auch mal delirierend?) zu Entscheidungen kommt. Oder ist das Bewußtsein von Freiheit ein konstituierendes Merkmal nur von *freien* Deliberationen? Wenn ja, wäre zu ventilieren, wann Deliberationen frei sind und wann nicht, und man sollte nicht glauben, dass diese Frage leicht zu beantworten ist. Es kann sein, dass sie gar nicht zu beantworten ist, weil sie nämlich möglicherweise gar keinen Sinn macht. Denn ein wenig Introspektion legt nahe, dass die Überlegungen zur Frage, ob man sich ein neues Auto kaufen soll oder nicht, nicht notwendig mit dem Erleben von Freiheit einhergehen, sondern eher mit einem Pendeln zwischen der Sorge um den CO₂-Abdruck einerseits und der Sorge um das Bankkonto andererseits. Spezieller kann man fragen, ob Deduktionen "freie" Prozesse sind, sofern man sich an die Regeln der Logik hält. Untersucht man das Problemlöseverhalten von Menschen, so geht es bei der Charakterisierung der dabei ablaufenden Prozesse im Allgemeinen nicht um die Freiheit einzelner Entscheidungen zwischen speziellen (Denk-)Operationen, sondern eben um die Struktur der involvierten Prozesse, die *bestimmt*, welche Operation in jedem Teilschritt des Lösungsprozesses gewählt wird, – wobei zufällige Entscheidungen nicht ausgeschlossen sind.

Ebenfalls findet man in Habermas' Ausführungen keine Begründungen für seine Behauptungen, dass Entscheidungen *nicht* vom Gehirn der Menschen getroffen werden; die Vorstellung, dass sein Gehirn und nicht er selbst Entscheidungen trifft empfindet er ja als Provokation⁴³. Dabei wäre es interessant, zu diskutieren, wie das Gehirn es anstellt, ein "Ich" zu generieren und welche Rolle das Ich bei Entscheidungen spielt. Die habermasschen Aussagen differenzieren schlicht nicht zwischen den Thesen des freien und des unfreien Willens. In seiner Kritik der habermasschen Arbeit hat Roth (2005) explizit gezeigt, wie Kognitionen, mithin auch Entscheidungsprozesse, durch Bezug auf die Ergebnisse der Hirnforschung erklärt werden können. Eine ausführliche philosophische Satz-für-Satz-Kritik der habermasschen Thesen findet man in Albert (2009).

nicht lesen, weil sie nur Trivialitäten sagen, in der Sprache des Hexeneinmaleins." Verg. Wissenschaftstheorie III (2), Abschnitt 4.5. Poppers Kritik gilt auch hier.

⁴³"Die objektivierende Sprache der Neurobiologie mutet dem "Gehirn" die grammatische Rolle zu, die bisher das "Ich" gespielt hat, aber sie findet damit an die Sprache der Alltagspsychologie keinen Anschluss. Die Provokation, die darin besteht, dass "das Gehirn" statt meiner "selbst" denken und handeln soll, ist gewiss nur eine grammatische Tatsache; aber so schirmt sich die Lebenswelt gegen kognitive Dissonanzen erfolgreich ab." (p. 872)

1.2.3.3 Julian Nida-Rümelin Alle folgenden Nida-Rümelin-Zitate beziehen sich auf diese Schrift. Nida-Rümelin stellt wie Dretske die Frage nach dem Verhältnis zwischen Freiheit und Kausalität (p. 2), d.h. wie sich die seiner Ansicht nach notwendig freie Entscheidung mit der "umfassenden kausalen Erklärbarkeit aller Ereignisse" in Einklang bringen läßt. Dazu erklärt er, dass "... nach meinem Verständnis Gründe eine zentrale Rolle für unseren Freiheitsbegriff spielen." und Entscheidungen finden nicht im Kopf, also nicht im Gehirn statt⁴⁴. Für Nida-Rümelin ist der Bezug auf Gründe konstitutiv für seinen Begriff des *Commonsense-Dualismus*, der durch den "Gegensatz zwischen dem von der Natur Gegebenen und menschlicher Setzung" charakterisiert wird; diese Setzung – gemeint sind wohl soziale Normen und Regeln – bezieht sich auf die Kultur, soll aber nicht mit Kultur gleichgesetzt werden. Der Commonsense-Dualismus führe auf eine *humanistische Perspektive*, die eine spezielle Definition von Freiheit erfordere: Nida-Rümelin will, wie er schreibt, erklären, "was Freiheit eigentlich ist" (p. 3): die "eigentliche Definition" der Freiheit sei, so Nida-Rümelin, die Annahme eines Spektrums,

"dessen eines Ende durch die "Willkürfreiheit" definiert ist, und dessen anderes Ende "jene Form von Freiheit [ist], die zurückgeht auf grundlegende, langfristig wirksame und möglicherweise schon lange vor der Ausführung einer konkreten Handlung getroffene Festlegungen einer Person." (p. 15),
Z 19

Diese Definition ist nicht mit der eingangs gegebenen Standarddefinition des Begriffs 'freier Wille' als Ausdruck für Entscheidungen ohne innere und äußere Zwänge kompatibel; hier geht Freiheit zurück auf "langfristig wirksame, möglicherweise schon lange vor der Ausführung einer konkreten Handlung getroffene Festlegungen einer Person"; bei Nida-Rümelin wird eine Theorie – die der humanistischen Perspektive – nicht in Abhängigkeit von der Definition von Willensfreiheit entwickelt, sondern umgekehrt die Definition nach Maßgabe der Theorie formuliert. Willkürfreiheit bestehe jedenfalls darin, sich im Falle völliger Indifferenz in Bezug auf die Entscheidungsalternativen eben rein willkürlich (soll wohl heißen: zufällig – wie auch immer das geschieht) für die eine oder andere Alternative zu entscheiden. Das andere Ende der Skala repräsentiert die Form von Freiheit, die Entscheidungen auf lange Deliberationen, Grundeinstellungen etc zurückführt. Diese Entscheidungen seien "lebensstrukturierende oder existentielle Entscheidungen" (p. 15). Dass einige unserer Entscheidungen lebensstrukturierend oder existentiell sein können, wird niemand bestreiten, bestreiten kann man allerdings, dass Entscheidungen durch Deliberationen zu freien Entscheidungen werden: durch Deliberationen werden möglicherweise Entscheidungen für bestimmte Alternativen ausgeschlossen, andere kommen in eine engere

⁴⁴In einem im Frankfurter Rundschau Magazin publizierten Streitgespräch am 03. 04. 2004 fragt der Philosoph Nida-Rümelin den Hirnforscher Wolf Singer, ob Singers Frau nicht mit ihm, Singer, schimpfe, wenn er ihr seine "berüchtigten Forschungsergebnisse präsentiere", denen zufolge der freie Wille eine Illusion sei, und ob es ihm, Singer, denn lieber sei, "wenn Ihr Gehirn Ihnen sagt, was Sie zu tun haben?".

Wahl, in die sie nicht gekommen wären, wären frühere Entscheidungen und Erfahrungen andere gewesen. Man könnte sagen, dass Nida-Rümelins Freiheit eigentlich gar keine ist, sondern dass hier eher von Unfreiheit die Rede ist, weil (i) die Deliberationen die Entscheidungen festlegen, und (ii) über den Prozess der Deliberation selbst nichts weiter ausgesagt wird. Es wird anscheinend angenommen, dass dieser Prozess "frei" abläuft, also wiederum nach dem Willen des Deliberierenden, aber dafür liefert Nida-Rümelin keinerlei Argumente. Wie bei Habermas wird der Deliberierungsprozess selbst gar nicht analysiert. Kritiker der These, der Wille sei frei, verweisen nicht zu Unrecht auf den Sachverhalt, dass die Details der Deliberationen der entscheidenden Person gar nicht bewußt sind und auf diese Weise Zwänge entstehen, die die Entscheidung am Ende unfrei sein läßt (z. B. Harris (2012), Strawson (1998)). Nida-Rümelin liefert keine Analyse der Möglichkeit eines freien Willens, sondern einfach eine Umdeutung des Begriffs des freien Willens. Im ersten Ansatz des einleitenden Abschnitts 1.1 wurde davon ausgegangen, dass das Prädikat 'frei' in Bezug auf Entscheidungen die Abwesenheit von Zwängen meint, und bisher ist Martin Heisenberg der einzige, der diesen Aspekt des Freiheitsbegriffs ernst genommen hat, vergl. Abschnitt 1.1.1. Ob sein Ansatz überzeugend ist, ist eine andere Frage. Der nächste offene Punkt ist, woher sich Nida-Rümelin die Freiheit nimmt, die Definition Z 19 zur "eigentlichen Definition" der Freiheit zu erheben, denn sie ist "eigentlich" völlig willkürlich. Nida-Rümelin liefert kein Argument, aus dem hervorgeht, dass man sein Konzept der 'eigentlichen Freiheit' anderen Konzeptionen vorzuziehen sei. Wie Dretske und Habermas ist Nida-Rümelin der Ansicht, dass kognitive Aktivitäten wie das Deliberieren und Konzipieren von Gründen gar nicht im Gehirn stattfinden (vergl. Fußnote 44, Seite 41). Dann aber ergibt sich die unbeantwortet bleibende Frage, wo denn dann die zu Entscheidungen führenden kognitiven Prozesse ablaufen. Insbesondere schreibt Nida-Rümelin, dass ein neurobiologischer Ansatz wegen der "naturalistischen Unterbestimmtheit von Gründen" inadäquat sei, weil:

"Gründe – beziehungsweise das Abwägen und Akzeptieren der letztlich besten Gründe – tatsächlich relevant sind, für das was ich tue. Wenn diese These gilt, so kann man Handlungen ohne Bezugnahme auf Gründe nicht erklären und erst recht nicht prognostizieren" (p. 17), Z 20

Diese These sei, so Nida-Rümelin, "mit einer bestimmten Form von Determinismus, nämlich dem in seiner naturalistischen Form, unvereinbar"; es wäre für die Diskussion nützlich, wenn Nida-Rümelin spezifizieren würde, was genau er mit "Determinismus in seiner naturalistischen Form" meint. Jedenfalls müsse man anerkennen, "dass Gründe für das menschliche Leben, die menschliche Verständigungspraxis, die menschliche Interaktion etc. eine irreduzible Rolle spielen."

Liest man die Aussagen Nida-Rümelins als umgangssprachliche Aussagen, so wird man wenig einzuwenden haben: wollen wir eine Erklärung für ein gegebenes Verhalten haben, so suchen wir nach Gründen dafür, und wenn wir keine Gründe

für dieses Verhalten finden können (Nida-Rümelin: "so kann man Handlungen ohne Bezugnahme auf Gründe nicht erklären"), so werden wir – trivialerweise – nicht erwarten, dass wir dieses Verhalten hätten vorhersagen können.

Nida-Rümelin will aber mehr als nur umgangssprachliche Betrachtungen anstellen, er möchte begründen, dass wir über eine Art von freiem Willen verfügen. Dazu wird der Begriff des Grundes nicht mehr nur umgangssprachlich verwendet, wobei die Begriffe 'Grund' und 'Ursache' praktisch denselben Umfang haben, sondern 'Grund' und 'Ursache' werden unterschiedlich definiert. 'Gründe' sind mentale Einheiten, 'Ursachen' sind neuronale Prozesse ("naturalistische Einheiten"). Die Steuerung des Verhaltens durch Gründe wird im Rahmen der Theorie Nida-Rümelins zu einer notwendigen, wenn auch nicht hinreichenden Voraussetzung für Prognosen. Eine Möglichkeit, die Aussagen Nida-Rümelins auf den Punkt zu bringen, ist diese:

Nach Zitat 20 sind Gründe für Handlungen relevant, und es soll gelten:

1. Eine Entscheidung geht entweder auf Gründe oder auf Ursachen zurück,
2. Wird eine Handlung (z.B. eine Entscheidung) ohne Bezug auf Gründe vorgenommen, so ist sie nicht erklärbar und nicht prognostizierbar.
3. Folgerung: gilt "Eine Handlung wird ohne Bezug auf Gründe vorgenommen", so folgt einerseits (s. 1.), dass sie das Resultat einer Ursache ist, andererseits nach dem vorangehenden Satz 2., dass sie nicht prognostizierbar ist. Aber Ursachen gehören nach Nida-Rümelin zur naturalistischen Perspektive⁴⁵, und damit zum Regime der angeblich deterministischen Naturgesetzmäßigkeit, weshalb die Handlung nun prognostizierbar wird⁴⁶. Die Struktur der von Nida-Rümelin spezifizierten Begriffe erweist sich demnach als opaque⁴⁷.

Nida-Rümelin wird derartigen Betrachtungen vermutlich nicht zustimmen wollen und sie als unphilosophische Pedanterien zurückweisen. Der Punkt ist, dass die Vagheit der Begriffe 'Grund' und 'Ursache', die doch die Basis für Nida-Rümelins philosophischen Ansatz bilden, in der Tat verstörenden Inkonsistenzen impliziert, und das wiederum nicht ohne Grund: die Unterscheidung von "naturalistischen" und mentalen, "nicht-naturalistischen" Systemen ist angesichts der in Abschnitt 1.1.1 noch einmal angedeuteten problematischen Natur des Kausalitäts- und Determinismusbegriffs kontraproduktiv, – Determinismus bedeutet nicht notwendig auch Prognostizierbarkeit und ist darüber hinaus selbst nicht begründbar (s. van Kampens Theorem). Auch ist unklar, warum der Begriff der Erklärung an den des Grundes gekoppelt wird, denn Erklärungen können doch auch durch Verweis auf Ursachen zustandekommen. Darüber hinaus sagt Nida-Rümelin nichts über den Deliberationsprozess: ist der Wille in welchem

⁴⁵Perspektiven: "... die humanistische, die auf Gründe, sowie die naturalistische, die auf Realursachen fokussiert ist". Nida-Rümelin (2007), p. 3 unten.

⁴⁶im Rahmen seiner Kritik des kantschen Ansatzes schreibt er: "Zum einen besteht die Möglichkeit einer naturwissenschaftlichen, kausalen Analyse, die den von außen objektiv beobachtbaren Verhaltensprozess einer Person ohne Rest prognostizieren kann", p. 6, d.h. 'kausal' impliziert 'prognostizierbar'.

⁴⁷'opaque' ist hier eine eher behutsame Umschreibung für 'widersprüchlich'.

Sinne auch immer noch frei, wenn der dieser Prozess bis auf stochastische Effekte deterministisch verläuft? Zufällige Entscheidungen sind eigentlich keine freien Entscheidungen. Wenn sich der Prozess als eine Folge von Abwägungen von Kosten- und Nutzenbewertungen darstellen läßt, so ist er bis auf stochastische Effekte bei den Bewertungen deterministisch⁴⁸. Ob der Begriff der humanistischen Perspektive hier noch eine klärende Funktion hat ist nicht offensichtlich.

Zur Illustration sei an die Mitte des 19-ten Jahrhunderts durch den Physiologen E.H. Weber (1795–1878) und den Mediziner G. T. Fechner (1801 – 1887) entwickelte Psychophysik erinnert, bei der es zunächst um die Beziehung zwischen Stimulusintensität und dazu korrespondierender Empfindungsstärke ging. Weber bestimmte für eine jeweils gegebene Stimulusintensität I einen Zuwachs ΔI ⁴⁹ der Intensität derart, dass die Intensitäten I und $I + \Delta I$ "gerade eben" unterschieden werden könnten. Dabei fand er, dass es bei keiner Sinnesmodalität und gegebener Intensität I ein scharf definiertes Inkrement ΔI_0 gibt derart, dass für $\Delta I < \Delta I_0$ die Intensitäten I und $I + \Delta I$ nicht unterschieden werden können, für $\Delta I \geq \Delta I_0$ aber I und $I + \Delta I$ stets voneinander unterschieden werden können. Vielmehr ist es so, dass für "hinreichend kleine" ΔI der Unterschied zwischen I und $I + \Delta I$ mit einer von I und ΔI abhängenden Wahrscheinlichkeit $p(I, \Delta I)$ entdeckt wird, wobei $p(I, \Delta I)$ monoton mit ΔI wächst, d.h. gegen 1 strebt. Die Entdeckung eines Unterschieds zwischen I und ΔI ist also kein deterministischer, sondern ein probabilistischer Prozess. Für einen gegebenen Versuchsdurchgang (der Versuchsperson (Vp) wird ein Stimulus s_1 mit der Intensität I und ein zweiter Stimulus s_2 mit der Intensität $I + \Delta I$ gezeigt, wobei entweder $\Delta I = 0$ oder $\Delta I > 0$, aber hinreichend klein ist), und die Vp muß anzeigen, ob $s_1 = s_2$ ($\Delta I = 0$) oder $s_1 \neq s_2$ ($\Delta I > 0$) ist. Die Antwort der Vp ist nicht prognostizierbar in dem Sinne, dass nicht mit Gewissheit vorausgesagt werden kann, welche Antwort die Vp geben wird, obwohl wir es doch mit "Realursachen", d.h. mit neurobiologischen Zuständen zu tun haben. Der Grund für diese Indeterminiertheit ist die grundsätzliche Stochastizität der neuronalen Prozesse, mit der Stimuli mit den Intensitäten I und $I + \Delta I$ verarbeitet werden. Diese Stochastizität gilt allgemein im Gehirn und führt auch bei Kognitionen "höherer Ordnung" zu Zufälligkeiten, die u.a. zu Aufmerksamkeitsschwankungen, zu Denkfehlern etc. führen können. In der empirischen Kognitionsforschung werden derartige Prozesse durchaus untersucht, bei den philosophischen Betrachtungen spielen sie so gut wie keine Rolle.

Die beschriebene Unsicherheit gilt also auch für komplexere Entscheidun-

⁴⁸Die Statistik hält für Entscheidungen unter Unsicherheit, bei denen die "Outcomes" nur mit Wahrscheinlichkeiten kleiner als 1 eintreten eine Reihe von Größen wie relatives Risiko, Odds Ratio etc bereit, anhand derer Entscheidungen in deterministischer Manier getroffen werden können. Diese müssen einer konkreten Person gar nicht bekannt sein, sie hat möglicherweise ganz ideosynkratische Entscheidungsregeln, die auf einen deterministischen Entscheidungsprozess hinaus laufen.

⁴⁹ Δ ist der griechische Buchstabe Delta, er wird oft zur Bezeichnung eines Unterschiedes (Differenz) verwendet.

gen, wenn die Folgen einer Entscheidung nicht mit Sicherheit überblickt werden können: Prognosen gelten immer nur mit bestimmten Wahrscheinlichkeiten, man denke an die Entscheidungen von Experten über das zu erwartende Wirtschaftswachstum des kommenden Jahres, oder die Entscheidung, eine bestimmte Investition zu tätigen oder nicht: ob die *richtige* Entscheidung getroffen wird ist i. A. nicht mit Sicherheit prognostizierbar. Dann würde nach Nida-Rümelin folgen, dass derlei Entscheidungen nicht in Bezug auf Gründe zustandekommen und demnach unfrei sind. Dann fragt sich aber, ob man überhaupt von freien Entscheidungen sprechen kann, sofern sie von komplexerer Natur sind, und die ganze Argumentation von Nida-Rümelin kollabiert. Wenn Nida-Rümelinschreibt, man müsse anerkennen, "dass Gründe für das menschliche Leben, die menschliche Verständigungspraxis, die menschliche Interaktion etc. eine irreduzible Rolle spielen," so kann man ihm zustimmen, so lange man nicht nach der genauen Bedeutung von 'irreduzibel' fragt, und sofern man sich an die umgangssprachliche Interpretation des Ausdrucks 'Grund' hält, womit man aber den Boden der Philosophie Nida-Rümelins verläßt. Es stellt sich die Frage, welchen Sinn nun seine Aussage über die "naturalistische Unterbestimmtheit von Gründen" hat. Die bloße Behauptung dieser Unterbestimmtheit ist ja kein Beweis für die Ungültigkeit der Gegenthese, dass Gründe ebenfalls naturalistisch bestimmt sind, zumal Nida-Rümelins Begriff der Unterbestimmtheit im hier gegebenen Zusammenhang selbst unterbestimmt ist. Die Gegenthese hat wiederum den Vorteil, ohne das Postulat, Entscheidungen fänden nicht im Gehirn statt, auszukommen. Warum ist es humanistisch, die biologische Basis unserer Kognitionen zu bestreiten? Nida-Rümelin suggeriert, weil wir uns mit der Akzeptanz einer biologischen Basis den "Determinismus in seiner naturalistischen Form" einhandeln, der freie Entscheidungen unmöglich mache. Aber wie die Diskussionen über den Determinismus gezeigt haben, ist dieser keine notwendige Implikation des naturalistischen Ansatzes, der nicht weniger humanistisch ist als Nida-Rümelins. Wer sich über die inhumanen Implikationen des humanistischen Bekenntnisses zum freien Willen informieren möchte, werfe einen Blick in das Kapitel 14 (The joy of punishment) von Sapolsky (2023).

1.2.3.4 Alva Noë Der amerikanische Philosoph Alva Noë (1964 –) publizierte 2009/2010 ein Buch mit dem Titel *Du bist nicht Dein Gehirn – Eine radikale Philosophie des Bewußtseins*. Die Seitenzahlen für die folgenden Zitate beziehen sich auf die deutsche Ausgabe. Noë geht nicht explizit auf die Frage nach dem freien Willen ein, sein zentrales Thema ist die Frage, ob die Aktivitäten des Gehirns die Basis des Bewußtsein bilden. Sollten sie es sein, würde sich die Frage nach dem freien Willen stellen, weil dann ja materielle Abläufe das Bewußtsein bestimmen und der Wille dann möglicherweise nicht frei ist. Ist das Bewußtsein aber unabhängig vom Gehirn, so sind Bewußtsein und damit bewußtes Handeln unabhängig von den kausalen Abhängigkeiten, die materielle Prozesse bestimmen, und damit kann dann auch der Wille frei sein. Betrachten wir also Noës Argumentationen.

Noë hat herausgefunden, wie wir die Antwort auf die Frage nach dem Bewußtsein finden können.

Um das Bewusstsein besser zu verstehen, müssen wir Abstand nehmen von dem nach innen gerichteten, neuronalen Mikrofokus ... Das Bewusstsein ist in den dynamischen Lebensprozess des gesamten, in eine Umwelt eingebetteten Lebewesens angesiedelt. Und nur wenn wir diese ganzheitliche Sicht auf die Lebensaktivitäten eines Menschen oder Tieres einnehmen, können wir allmählich verstehen, was das Gehirn zu unserem bewussten Erleben beiträgt.

Dies ist ein positives Buch. die menschliche Erfahrung ist ein Tanz, der sich in der Welt entfaltet. Wir sind nicht unser Gehirn. Wir sind nicht im Gefängnis unserer Vorstellungen und Empfindungen gefangen. Das Phänomen des Bewusstseins ist ebenso wie das des Lebens ein mit der Welt verwobener, dynamischer Prozess. Wir sind zuhause in der Welt, die uns umgibt. Wir stecken nicht in unserem Kopf." (p. 13), Z 21

Auf Seite 19 schreibt er, dass er den Leser

"... überzeugen[will], dass die Grundthese der Bewußtseinsforschung grundsätzlich verfehlt ist. Das Bewusstsein entsteht nicht im Gehirn. Und daran liegt es auch, dass bis jetzt niemand eine überzeugende Erklärung aus den neuronalen Grundlagen liefern konnte." (p. 19), Z 22

und für den Geist gilt nach Noë :

"... der Geist scheint ein Paradoxon in sich zu tragen: Er ist Teil unserer Natur, und doch eignet er sich nicht zum Gegenstand der Naturwissenschaft." (p. 56), Z 23

Die Meinung, teilt er mit vielen Geisteswissenschaftlern. Wir brauchen seiner Ansicht nach eine "Perspektive, aus der sich der bedeutungstragende, nicht-mechanische Charakter bewussten Lebens überhaupt erst zeigt" (p. 56). So bewegen sich Bakterien mit ihren Flagellen gezielt auf Zuckerkonzentrationen zu, wobei sie "maschinenartig mit ihrer Umwelt verzahnt erscheinen. Indem wir die Bakterie derart beschreiben, sehen wir sie aber eigentlich nicht mehr als rein mechanisches und rein physikalisches Wesen, sondern als Einheit, deren Handlungen tatsächlich als *Handlungen* gewertet werden können" (Kursivsetzung durch Noë). Sie wird von der Zuckerkonzentration angezogen, weil sie Zucker zum Leben benötigt, und "das heißt, dass es in seiner Umgebung in einer Beziehung steht." (p. 57).

Noë fragt nun, "wie aus dem unverkennbaren Knacken, Knistern und Knallen der Neurone eine Erfahrung [wird], die eine bestimmte Beschaffenheit und nicht eine andere hat?" (p. 71-72), – wir hören etwas, oder wir sehen etwas. Noë versichert noch einmal, dass die Wissenschaft diese Frage nicht beantworten könne, bis zum heutigen Tag seien wir nicht in der Lage, die sogenannte Erklärungslücke (explanatory gap) zu schließen. Das sei auch nicht verwunderlich:

Wir können die qualitative Beschaffenheit der Erfahrung nicht mit der intrinsischen Natur der Hirnvorgänge erklären, weil es an Hirnvorgängen nichts gibt, was beispielsweise besonders "visuell" wäre." (p. 72), Z 24

Würde man Noë sagen, dass jeder Augenarzt ihm sagen könnte, welche Schäden im visuellen System zu welchen Wahrnehmungsdefiziten führen, so würde ihn das vermutlich wenig beeindrucken, denn für ihn impliziert das Zitat 24, dass dem Organismus gleichzeitig "Geist" zugeordnet werden kann. Beim Bakterium ist das ein eher rudimentärer Geist, aber es ist Geist. Denn "die Frage nach dem Geist ist die Frage nach dem Leben" (p. 58). Die Bewegungen des Bakteriums bedeuten die Existenz von Handlungskompetenz, und damit habe es "ein zumindest rudimentäres Bewußtsein". Die Frage nach dem Bewußtsein sei die Frage nach dem Leben, und es gehe darum, wie Leben in der Natur entsteht (p. 59). Das Leben einer Bakterie liege nicht in ihrem Inneren verborgen, es sei vielmehr ein Kräftespiel, "an dem die in eine Umgebung eingebettete Bakterie selbst teilnimmt. Und das Gleiche gilt auch für das Bewußtsein in einem weiteren Sinne." (p.60).

Man wird protestieren wollen, denn nach allem, was wir über Bakterien wissen, haben Bakterien keine Neurone und damit auch kein Gehirn. Dieses Argument zieht bei Noë nicht, weil es seiner Ansicht nach, wenn es um das Bewusstsein geht, ja gar nicht auf die Neurone ankommt. Schließlich regeln Neurone auch die Darmtätigkeit, die uns im Allgemeinen nicht bewusst ist. Andererseits enthält das Gehirn höherer Lebenwesen sowohl Seh- wie Hörzentren, – wenn es auf die Neurone dieser Zentren gar nicht ankommt – wozu gibt es diese Zentren dann? Man könnte annehmen, dass die Sinneszentren so etwas wie Sende- und Empfangszentren für die extracraniale Geisteswelt ist. Wissenschaftstheoretisch ist ein solcher Ansatz wenig attraktiv, weil Noë einfach behauptet, das "geistige" Aktivität deswegen keine neuronale Basis hat, weil man bis jetzt nicht weiß, wie neuronale Aktivität Bewußtsein erzeugt. Aber aus diesem Nichtwissen folgt ja nicht, dass es da nichts zu wissen gibt. Mit der Verlagerung psychischer Aktivität in eine rein psychische Welt außerhalb des Kopfes kehrt Noë einfach zur kantschen Zweiten Welt der Noumena zurück, von der man auch nichts weiter weiß als dass Kant sie postuliert hat.

Man fragt sich auch, wozu der Mensch neuronale Sprachzentren hat. Primaten ohne derartige Zentren können keine Sprache auf humanem Niveau erlernen. Noë scheint hier kein Problem zu sehen: Sprache sei u.a. eine soziale Praxis innerhalb einer Sprachgemeinschaft, deshalb seien unsere kognitiven Fähigkeiten auf eine *soziolinguistische Umwelt* angewiesen: "Unsere Geister überqueren die Grenze des Schädels und werden von einem gemeinsamen soziolinguistischen Gerüst getragen" (p. 109). So kommt er zu der Feststellung

"Die Bedeutung ist nicht im Kopf" (p. 110), Z 25

Noë hat recht mit der Aussage, dass die Neurowissenschaft in den letzten Jahrzehnten in Bezug auf die Frage nach der Entstehung des Bewußtseins aus der

neuronalen Aktivität nicht so recht vorangekommen ist. Wie oben schon gesagt wurde folgt daraus noch lange nicht, dass es keine neurowissenschaftliche Erklärung geben kann, weil alles Psychische eben gar nicht im Kopf abläuft. Man hat weit über 2000 Jahre gerätselt, was Gravitation "eigentlich" ist, bis Einstein eine akzeptable Theorie lieferte. Der Große Fermatsche Satz⁵⁰ konnte mehr als 350 Jahre nicht bewiesen werden, obwohl die jeweils besten Mathematiker versucht hatten, einen Beweis zu finden, – man hätte glauben können, dass kein Beweis existiert. Bis der britische Mathematiker Andrew Wiles 1993 einen Beweis vorlegte; dieser Beweis enthielt noch kleine Fehler, eine fehlerfreie Version des Beweises liegt seit 1995 vor. Die Philosophie des Enaktivismus liefert ebenfalls keine Antwort auf die Frage, wie das Bewußtsein entsteht. Dass wir mit der Umwelt interagieren ist eine Einsicht ohne jeden Neuigkeitswert, die Aussage des Zitats Z = 25 ist völlig unbewiesen, Noë kann auf keine Daten verweisen, die die Hypothese der Repräsentation der Umwelt im Gehirn in Frage stellen, und ebenso wenig liefert er theoretische Argumente, sondern nur suggestive Wiederholungen von bestenfalls Plausibilitätsbetrachtungen unter Vernachlässigung aller Evidenz für die Rolle neuronaler Systeme, die allein im medizinischen Bereich akkumuliert wurde. Noë behauptet, die klassische Vorstellung, dass das Nachdenken und Argumentieren etc im Kopf stattfindet, sei von der Philosophie der letzten sechzig Jahre "vollkommen widerlegt" worden (p. 111). Derartige Aussagen verblüffen eher als dass sie zu einem Aha-Erlebnis führen, insbesondere wenn er zur Illustration auf die in der Neuro- und Kognitionswissenschaft berühmt gewordenen Arbeiten von David H. Hubel (1926 – 2013) und Thorsten N. Wiesel (1924 –) zur Struktur des visuellen Kortex von Katzen und Primaten eingeht, für die die beiden Forscher 1981 mit dem Nobelpreis für Medizin geehrt wurden. Noë erkennt die Leistung der beiden Forscher insofern an, als sie "fleissig gemessen" haben, aber er bezweifelt den Wert der Arbeit für das Verständnis der Prozesse der visuellen Wahrnehmung: der Blick auf die neuronalen Aktivitäten im visuellen Kortex sagen seiner Ansicht nach so gut wie nichts über deren Rolle im Wahrnehmungsprozess aus. Dabei wäre es doch nach des Philosophen Ritters Meinung die vornehmste Aufgabe von Philosophen wie Noë, uns zu zeigen, welche Bedeutung die Ergebnisse von Hubel & Wiesel nun wirklich haben (vergl. Zitat 2, Seite 3).

Noës Theorie wurde in den Feuilletons sehr positiv aufgenommen⁵¹, wird aber auch in Philosophenkreisen durchaus geschätzt, wenn auch störende Vagheiten in seinen Ausführungen kritisiert werden⁵². Nüchterner ist das Fazit einer im *Scientific American* erschienenen Besprechung von Noës Buch:

"One comes away from the book without a definitive example of a conscious state that would require more than a brain."⁵³, Z 26

⁵⁰Die Aussage $a^n + b^n = c^n$ mit $a, b, c, n \in \mathbb{N}$ gilt nur für $n \leq 2$

⁵¹So schreibt z.B. Helmut Mayer in der FAZ vom 21. 12. 2010: "Alva Noë führt exzellent vor Augen, warum sich in den Neuronen nicht finden lässt, was Bewusstsein ausmacht".

⁵²Eine ausführliche Besprechung liefert Rachlin, H. (2012)

⁵³J. Hoffman: MIND Reviews: Out of Our Heads, by Alva Noë, April 2009 – Recommenda-

Wir wissen zwar nicht, wie das Gehirn Bewußtsein erzeugt, aber dass es ohne Gehirn nicht erzeugt wird kann kaum geleugnet werden, dafür sprechen allein schon die medizinischen Befunde zu kognitiven Ausfällen nach Unfällen oder Erkrankungen.

1.2.4 Max Plancks Betrachtungen

Die Beziehung zwischen einer deterministischen Auffassung des Naturgeschehens und der These, der Wille sei frei, ist 1932 von Max Planck (Planck (1932) sehr genau analysiert worden, – mit einer überraschenden Schlußfolgerung.

Zunächst stellt Planck fest, dass das allgemeine Kausalgesetz wegen seines transzendentalen Charakters wahr sei (p. 157); transzendental bedeutet hier so viel wie notwendig, also eine Bedingung der Möglichkeit, eine kausale Folge von Ereignissen als solche zu verstehen. Planck argumentiert nun, dass das Kausalgesetz nur in makroskopischen, nicht aber in mikroskopischen Betrachtungen anwendbar sei, wobei mit 'mikroskopisch' die Prozesse auf atomarem und molekularem Niveau gemeint sind, das nur der berühmten, von Laplace in einem Gedankenexperiment angeführten *intelligence* zugänglich ist, wenn sie den Kausalitäten der physikalischen Welt bis ins wirklich letzte Detail nachgehen will. Nun muß geklärt werden, ob es der später in 'Laplacescher Dämon' umbenannten *intelligence* möglich ist, Raum für einen freien Willen zu finden, denn physikalisch ist ja, nach Voraussetzung, alles determiniert (p. 159).

Planck überlegt, ob ein Punkt existiert, an dem jede kausale Gedankenfolge aufhört. Ein solcher Punkt könne wissenschaftlich nicht erreicht werden. es gebe ihn nur in der "unmessbaren Welt von Geist und Materie" ("in the immeasurable world of mind and matter") (p. 161), wo keine kausale Forschungsmethode mehr anwendbar sei (Planck). Dieser Punkt sei das individuelle Ego:

"It is a small point in the universal realm of being; but in itself it is a whole world, embracing our emotional life, our will and our thought." (p. 161),
Z 27

Planck befindet nun, dass es keinen Grund gebe, sich nicht zum Beobachter dessen, was in ihm geschieht, zu machen. Es sei logisch, wenn auch nicht praktisch, möglich, im Rahmen eines Gedankenexperiments zum Laplaceschen Dämon zu werden, also alle kausalen Abhängigkeiten zu erkennen, allerdings nur bis auf die eigenen. Die eigenen kausalen Abhängigkeiten könne man nicht erkennen, – denn kein Auge könne sich selbst sehen, so Plancks Illustration seines Arguments. Skeptiker des freien Willens können dieses Argument zur Basis eines Arguments gegen die These der Existenz eines freien Willens machen: die angebliche Freiheit des Ichs sei nur das Resultat eines Mangels an vollständigem Verstehen aller kausalen Relationen. Aber Planck weist diese Argumentation

tions from Scientific American MIND

sofort zurück: sie sei falsch, sie gleiche dem Argument, ein Läufer könne seinen eigenen Schatten nicht überholen, weil es ihm an Geschwindigkeit mangle. Die Wahrheit der Freiheit des Individuums resultiere aus der Logik:

"The fact that the individual here and now, in regard to his own living present act, cannot be subject to the law of causation is a truth that is based on a perfectly sound logical foundation of an a priori kind, such as the axiom that the part is never greater than the whole." (p. 163-164).

Z 28

Planck bezieht die Superintelligenz des Laplaceschen Dämons in seine Argumentation ein: auch der Dämon müsste die Aktivitäten seines eigenen Ichs in dem Moment, in dem er sein Ich beobachtet erkennen, was eben gerade nicht möglich sei (man erinnere sich an das Auge, das sich nicht selbst sehen kann);

"We cannot possibly study ourselves at the moment or within the environment of an given activity. Here is the place where the freedom of the will comes in and establishes itself, without usurping the right of any rival." (p. 164)

Z 29

Einige Zeilen weiter fügt Planck an, die Freiheit des Ichs und seine Unabhängigkeit von der Kausalkette resultiere aus dem Diktat des Bewußtseins (p. 165). Das "eiserne Gesetz der logischen Konsistenz" impliziere, dass wir die kausalen Grundlagen unserer eigenen Zukunft und deren Vorhersagbarkeit nicht bestimmen können (p. 167).

Planck hat sich viele Jahre mit der Frage des freien Willens beschäftigt. In einem Vortrag (1946) schreibt er:

"Von außen betrachtet ist der Wille kausal determiniert, von innen betrachtet ist der Wille frei. Mit der Feststellung dieses Sachverhaltes erledigt sich das Problem der Willensfreiheit. Es ist nur dadurch entstanden, dass man nicht darauf geachtet hat, den Standpunkt der Betrachtung ausdrücklich festzulegen und einzuhalten. Wir haben hier ein Musterbeispiel für ein Scheinproblem. Wenn diese Wahrheit gegenwärtig auch noch mehrfach bestritten wird, so besteht doch für mich kein Zweifel darüber, dass es nur eine Frage der Zeit ist, wann sie sich zur allgemeinen Anerkennung durchringen wird." (p. 9),

Z 30

Das Problem des freien Willens ein Scheinproblem! Zu der Vermutung, dass es sich um ein Scheinproblem handeln könnte, hatten schon die intuitiven Betrachtungen zum Begriff des freien Willens in Abschnitt 1.1 geführt. Gleichwohl ist nicht ganz klar, was Planck wirklich meint: bevor er zu der Aussage im Zitat 30 kommt, hat er ja festgestellt, dass die Entscheidungen im Ich verankert sind, das sich der Kausalität entzieht (Zitate 28 und 29), dass aber das Universum mit allem, was darin ist, gleichwohl determiniert sei. Planck wiederholt viele seiner Argumente aus den *Philosophical Essays* aus dem Jahr 1932. Das wirft die Frage

auf, wann ein Problem ein Scheinproblem ist (zum Begriff des Scheinproblems vergl. Wissenschaftstheorie III (1), Abschnitt 6.3). Plancks Schlußfolgerungen sind oft *ad hoc* und von der Überzeugung geleitet, dass moralische Verantwortung nicht ohne die Freiheit des Willens möglich sei. Dass die Aussage, dass ein Individuum im Hier und Jetzt nicht dem Kausalgesetz unterliegt (Zitat 28), aus rein logischen Gründen wahr sein muß, erscheint als Behauptung und wird nicht weiter begründet. Eine detaillierte Kritik seiner Argumentation überschreitet den Rahmen dieser kurzen Darstellung, eine ausführliche Diskussion findet man bei Griffel (1998). Griffel findet in Plancks Vorträgen über die Willensfreiheit eine Reihe argumentativer Sprünge, die erst vor dem Hintergrund der Gesamtphilosophie Plancks verständlicher werden (vergl. Scheibe (2006)⁵⁴). Plancks Aussage in Zitat 28 erinnert an ein Problem, das Singer (Zitat 12, Seite 29) auf den Punkt gebracht hat. Singers Aussage in Zitat 12 kann bedeuten, dass Planck mit seiner Deutung des "Ichs" recht hat⁵⁵. Darüber ist aber noch nicht endgültig entschieden worden, und man sollte nicht vorschnell aufgeben, – mehr Detailkenntnisse über die neuronale Basis kognitiver Prozesse sind dazu dringend erforderlich.

Planck war erklärter Determinist, und bei dieser Grundposition ist es viel schwieriger, für die Willensfreiheit als für die Unfreiheit des Willens zu argumentieren. Die Frage, "ob es einen Punkt gibt, an dem jede kausale Gedankenfolge aufhört" und die dazu korrespondierende Aussage in Zitat 27 auf Seite 49 illustrieren seine Bemühungen, im Reich eines allgemeinen Determinismus einen Ort für die Freiheit der Entscheidung zu finden. Planck schreibt, dass wir uns selbst und unsere Aktivität in einem gegebenen Moment nicht erkunden (study) können und hieraus der freie Wille entstehe, diese Aussage sei logisch zwingend (vergl. Zitat 28 oben). Mit der Aussage, eine Selbstbeobachtung für einen gegebenen Moment einer kognitiven Aktivität sei nicht möglich, mag er recht haben (vergl. Singers Aussage in Zitat 12, Seite 29) – aber dass daraus die Freiheit des Willens folgt ist nicht offenkundig. Planck macht keinen Versuch, seine Argumentation zu formalisieren, um ihr Stringenz zu verleihen. Vermutlich hat Planck Alan Turings Arbeit aus dem Jahr 1936 nicht gekannt; sie liefert, wie im folgenden Abschnitt beschrieben wird, eine Basis für eine strengere Diskussion der Planckschen Betrachtungen.

1.2.5 Seth Lloyds rekursive Argumentation

Tatsächlich ist ein solcher Beweis von dem Physiker und Informatiker Seth Lloyd (2012) vorgelegt worden; eine Skizze des Beweises findet man im Anhang. Hier

⁵⁴Einen guten Überblick findet man in

<https://www.informationphilosopher.com/solutions/scientists/planck/>.

⁵⁵Die Anwendung der Theoreme von Gödel und Turing auf kontinuierliche Systeme wird gelegentlich kritisiert, weil die Theoreme sich auf diskrete Systeme beziehen. Neuere Arbeiten legen nahe, dass die Theoreme auf kontinuierliche Systeme übertragbar sind, z.B. Granicky, M.S. (1995), Vicente (2023)

soll nur die Aussage des Lloydschen Befundes diskutiert werden⁵⁶. Der Beweis basiert auf Turings (1936) Halte-Theorem, demzufolge es keinen Algorithmus T_2 gibt, mit dem festgestellt werden kann, ob ein gegebener Algorithmus T_1 für einen gegebenen Input stoppt oder in eine unendliche Schleife gerät; das Halte-Theorem wird im Anhang 2 vorgestellt.

Lloyd illustriert seine These zunächst informell. Er beschreibt sein Verhalten in einem Restaurant: er studiert die Speisekarte, ohne zu wissen, was er wählen wird. Seine Frau weiß es allerdings schon, weil er immer dasselbe Gericht wählt. Aber Lloyd ist nicht mit der Absicht ins Restaurant gegangen, genau dieses Gericht zu bestellen. Der Punkt ist, dass ihm sein Entscheidungsprozess nicht bewußt ist, – er weiß erst, was er möchte, nach einer gewissen Zeit des Studiums der Speisekarte. Er kann es nicht vor Ablauf des Entscheidungsprozesses wissen, denn könnte er es wissen, gäbe es einen einen zum Entscheidungsprozess parallel laufenden Prozess, der das Ergebnis des Entscheidungsprozesses vorhersagen könnte, – und dieser Prozess wäre dann der eigentliche Entscheidungsprozess, für den dann dasselbe Argument gelten würde. Jeder Prozess, der den Entscheidungsprozess "beobachtet", kann dem Entscheidungsprozess nur nachlaufen, andernfalls wäre der Beobachtungsprozess der Entscheidungsprozess. Deshalb kann der Beobachtungsprozess, sollte es ihn denn geben, den Entscheidungsprozess auch nicht beeinflussen. Moralische Aspekte, sollten sie für eine bestimmte Entscheidung von Bedeutung sein, werden beim Entscheidungsprozess bereits berücksichtigt.

Bei der Diskussion der Möglichkeit eines freien Willens geht es u.a. um die (Selbst-)Wahrnehmung und eventuelle Beeinflussung oder Steuerung des Prozesses, der zu einer Entscheidung führt. Aber die Beeinflussung bzw. Steuerung sind dann bereits Teil des Entscheidungsprozesses. Nimmt man wie viele Kompatibilisten an, dass diese Prozesse irgendwo außerhalb des Hirns ablaufen, so hat man die Diskussion von Entscheidungsprozessen kurz- und abgeschlossen, denn über den Ort und die Struktur derartiger extracranialer Abläufe ist nichts bekannt. Es wird nur *postuliert* und keineswegs bewiesen, dass diese Abläufe nicht durch die Fesseln des Kausalitätsprinzips eingengt werden. Die Behauptung, Entscheidungsprozesse seien keine Hirnprozesse, liefert keinerlei Erklärung, weil es völlig unklar ist, was man sich unter außerhalb des Gehirns ablaufenden Prozessen überhaupt vorstellen soll.

Geht man aber davon aus, dass Entscheidungsprozesse Hirnprozesse sind, so muß man akzeptieren, dass es sich letztlich um physikalische Prozesse handelt. Lloyd führt aus (p. 11), dass derartige Prozesse durch Turingmaschinen simuliert werden können. Auch Folgen von Argumentationen können, wenn sie durch Regeln gesteuert werden, in ihrem Kern durch Algorithmen und deswegen durch Turingmaschinen abgebildet und deshalb auch simuliert werden. Lloyd nennt derartige Argumentationen *rekursive Argumentationen*. Er argumentiert, dass

⁵⁶Interessierte Leser können die Arbeit direkt aus dem Netz herunterladen: Lloyd, S.: A Turing Test for Free Will. <https://arxiv.org/abs/1310.3225>

die genannten physikalischen Prozesse die rekursiven Argumentationen umfassen bzw. durch derartige Argumentationen repräsentiert werden können.

Die zentrale These ist, dass Entscheidungen für das entscheidende Subjekt nicht vorhersagbar sind: es gibt stets eine Deliberation, die abläuft, wenn entschieden werden muß. Dieser Prozess kann sehr kurz sein oder auch länglich sein, etwa wenn moralische Aspekte eine Rolle spielen. Für den Vegetarier benötigt die Entscheidung zwischen Wurst und Käse kaum Zeit, weil sie vorprogrammiert ist, während die Entscheidung, ob das neue Eigenheim mit Gas oder mittels Wärmepumpe beheizt werden soll länger dauern kann, weil unmittelbare ökonomische Aspekte einerseits und eher ethische Aspekte andererseits gegeneinander abgewogen werden müssen. Am Ende erscheint das Resultat des Prozesses im Bewußtsein als vom Ich erzeugt und *deshalb* als freie Entscheidung. Dabei ist es unerheblich, ob der Entscheidungsprozess deterministisch oder probabilistisch abgelaufen ist, jede während des Deliberationsprozesses getroffene Sub-Entscheidung ist "meine", von meinem "Ich" getroffene und deshalb "freie" Entscheidung. Dass die bei der Deliberation ablaufenden, durch Aufmerksamkeitsfluktuationen, ästhetische und/oder moralische Präferenzen etc gesteuerte, aber nicht notwendig bewußte Prozesse gar nicht "frei" sind geht in das Erlebnis von Freiheit gar nicht ein. Hier ist der von Planck erwähnte Punkt (Zitat 27, Seite 49 oben), in dem Planck die Freiheit der Entscheidung angenommen hat.... Es sieht aber aus, als habe der Begriff der Freiheit in diesem Zusammenhang keinen Sinn mehr.... es ist "meine" Entscheidung, denn ich bin so, wie ich nun mal bin.

Lloyds Ergebnisse lassen sich wie folgt zusammenfassen (den formalen Beweis findet man in Lloyd (2012), p. 10; er basiert wie der des Halte-Problems auf dem Cantorsche Diagonalverfahren (vergl. Anhang):

1. Es existiert keine allgemeine Methode, mit der sich die Frage beantworten ließe, ob eine entscheidende Person überhaupt in endlicher Zeit zu einer Entscheidung kommt; diese Aussage ist Turings Halte-Problem.
2. Wenn die entscheidende Person sich innerhalb eines endlichen Zeitintervalls entscheiden muß, so kann jede allgemeine Methode, mit der die Entscheidung der Person bestimmt ("vorhergesagt") werden kann, für diese Bestimmung mehr Zeit benötigen, als der Entscheidungsprozess der Person selbst.
3. Jeder durch eine universelle Turingmaschine repräsentierbare Entscheider kann nicht alle Fragen über sein zukünftiges Verhalten beantworten.
4. Jeder zeitlich begrenzte Entscheider benötigt mehr Zeit, den Entscheidungsprozess zu simulieren als ihn direkt durchzuführen.

Diese Ergebnisse erklären, warum die meisten Menschen sich als *frei* entscheidend erleben: die Entscheidung erscheint im Bewußtsein, die entscheidende Person kann sagen: "Ich habe die Entscheidung getroffen!", das "Ich" erscheint als

die entscheidende Instanz. Dies gilt unabhängig davon, ob der Entscheidungsprozess deterministisch oder probabilistisch verläuft. Wir können den Entscheidungsprozess nicht mit einem parallel zu diesem Prozess laufenden Prozess beobachten und dementsprechend auch nicht auf "freie Weise" beeinflussen. Dieser Befund ist vermutlich das, was Planck meinte, dass ein Auge sich nicht selbst sehen und ein Läufer seinen Schatten nicht einholen kann, was ihn aber noch nicht dazu brachte, die (Hypo-)These der Existenz des freien Willens aufzugeben, – eine Metapher ist noch kein Beweis. Entscheidet man sich dazu, die These der Existenz des freien Willens beizubehalten, so nimmt man eine Art impliziter Umdefinition des Begriffs des freien Willens vor: der Wille ist frei, weil "Ich" die Entscheidungen treffe (sofern sie nicht von außen erzwungen werden) und der Entscheidungsprozess von "mir" durchgeführt wird, – der Entscheidungsprozess ist gewissermaßen in den Prozess, der das Bewußtsein generiert, integriert. Dass diese Prozesse von meinem Gesamtzustand abhängen und insofern auch determiniert werden, geht aber in diese Definition nicht ein. Die Erfahrung der "Freiheit" des Willens ist sehr überzeugend – man kann von einem Evidenzerlebnis sprechen –, so dass es schwer, wenn nicht unmöglich wird, sozusagen in Eigenbeobachtung die Rolle der nicht bewußt ablaufenden Prozesse zu reflektieren.

1.2.6 Freier Wille und moralische Verantwortung

Die zentrale Motivation von Philosophen, die für die Freiheit des Willens argumentieren, scheint die Frage nach der moralischen Verantwortung zu sein. Ohne einen freien Willen könne es kein moralisches Handeln geben. Skeptiker des freien Willens sind der Ansicht, dass der freie Wille keine Voraussetzung für die Möglichkeit moralischen Handelns ist. Einigkeit herrscht aber über den Sachverhalt, dass wir, in habermasscher Ausdrucksweise, vergesellschaftete Individuen sind, d.h. wir haben Regeln – wenn auch nicht notwendig alle – des sozialen Handelns gelernt (Bandura & Walters, (1963)) haben wohl als erste eine empirisch untermauerte Theorie des sozialen Lernens präsentiert, der aber von Philosophen einerseits und Wissenschaftlern andererseits verschieden ausgelegt werden kann. So schreibt der Neurowissenschaftler W. Singer (2004):

"Genetische Faktoren, frühe Prägungen, soziale Lernvorgänge und aktuelle Auslöser, zu denen auch Befehle, Wünsche und Argumente anderer zählen, wirken stets untrennbar zusammen und legen das Ergebnis fest, gleich, ob sich Entscheidungen mehr unbewussten oder bewussten Motiven verdanken. Sie bestimmen gemeinsam die dynamischen Zustände der "entscheidenden" Nervennetze." (p. 254)

Z 31

Natürlich sind Philosophen wie Habermas, Nida-Rümelin, Noë etc mit dieser Interpretation empirischer Daten nicht einverstanden, aber Singer weist auch daraufhin, dass eine Straftat von den Gerichten nach Maßgabe der psychischen Befindlichkeit des Täters beurteilt wird: wird nämlich festgestellt, dass ein Tu-

mor die im Frontalhirn angesiedelten Kontrollfunktionen reduziert hat, so wird der Täter mit Nachsicht beurteilt:

”Der gleiche ”Defekt” kann aber auch unsichtbare neuronale Ursachen haben. Genetische Dispositionen können Verschaltungen hervorgebracht haben, die das Speichern oder Abrufen sozialer Regeln erschweren, oder die sozialen Regeln wurden nicht rechtzeitig und tief genug eingeprägt, oder es wurden von der Norm abweichende Regeln erlernt, oder die Fähigkeit zur rationalen Abwägung wurde wegen fehlgeleiteter Prägung ungenügend ausdifferenziert. Diese Liste ließe sich nahezu beliebig verlängern. Keiner kann anders, als er ist.” (Singer (2004), p. 254) Z 32

Zumindest bei den hier vorgestellten Philosophen werden derartige Sachverhalte noch nicht einmal erwähnt, die Idee, dass Entscheidungen nach Maßgabe von Gründen und außerhalb des Kopfes getroffen werden läßt alles, was mit dem Gehirn zusammenhängt, irrelevant erscheinen; wie Noë im Zusammenhang mit den Untersuchungen von Hubel & Wiesel argumentiert hat: die Philosophie und nicht die Naturwissenschaft beantwortet alle relevanten Fragen.

Vielleicht auch nicht. Der Hinweis auf die Notwendigkeit von psychologischen und medizinischen Therapien mag genügen, um derart abgehobene philosophische Theorien mit Zurückhaltung zu betrachten. Denn psychische Störungen verweisen ja auf die Tatsache, dass Menschen sich eben nicht frei gegen ihre Störung entscheiden können, auch wenn die Störung die Lebensqualität herabsetzt, – sie benötigen eben therapeutische Hilfe, und die ist nicht notwendig nur psychologischer, sondern gelegentlich auch pharmakologischer Art, d.h. sie wirkt auf physiologische Weise⁵⁷. Man fragt sich, welchen Erkenntniswert philosophische Theorien der hier vorgestellten Art überhaupt erzeugen. Der von Roth (2004) zitierte Philosoph J. Ritter, der der Ansicht war, dass Naturwissenschaftler die natürliche Welt zwar erforschen, aber nicht interpretieren können, weshalb sie die Interpretation besser den Philosophen überließen, verfügte offenbar über ein gesundes Selbstbewußtsein.

C.F. von Weizsäcker zu ’Heidegger: ”Die Philosophie billigt der Naturwissenschaft nur eine seinsvergessene Projektion der Wahrheit zu” (.v Weizsäcker (1988), p. 185) v. Weizsäckers Einschätzung der Einschätzung der Naturwissenschaft durch Heidegger⁵⁸

Es soll hier nichts gegen philosophische Reflexionen über empirische Befunde gesagt werden, sondern nur die Allgemeingültigkeit und der Ausschließlichkeitsanspruch philosophischer Betrachtungen in Frage gestellt werden. Philosophische Theorien sind keine formalen Theorien, wie man sie aus der Mathematik kennt und wo man aus Axiomen mit wohldefinierten Schlußregeln und

⁵⁷Habermas (2004) diskutiert die Rolle von Gründen und bedingter Freiheit im Unterschied zur Rolle von Ursachen ausführlich, allerdings am Ende nicht überzeugend, weil er nicht auf das, was bereits über die physische Basis bewußter Prozesse bekannt ist, eingeht.

⁵⁸v. Weizsäcker fragt weiter, wie man in der Philosophie verständlich macht, dass Naturwissenschaft überhaupt gelingt. Heideggers Antwort: ”Das weiß ich auch nicht.”

genau formulierten Definitionen "wahre" Aussagen deduziert. Eine Teilmenge von Aufsätzen aus der Philosophie des Geistes enthalten in erster Linie begriffliche Explikationen, gepaart mit apodiktischen Aussagen (z.B. Dretske: "Meaning itself is causally inert, powerless to initiate, ...", Zitat 10, Seite 29), und "Even if the things that have meaning are in the head, the meanings themselves aren't in the head." Zitat 11, Seite 29). Selbst wenn die Elaborationen der Annahmen strenger und weniger apodiktisch vorgenommen werden bleibt immer noch das Letztbegründungsproblem, nach dem man stets gezwungen ist, für Theorien jeweils bestimmte Annahmen machen zu müssen, die zuguterletzt nicht weiter begründbar und deshalb angreifbar sind⁵⁹. Da verschiedene Autoren nicht nur aus der Philosophie, sondern auch aus den Naturwissenschaften (man denke an die Physiker Anton Zeilinger (Zitat 6, 10 und Gerard t'Hooft (Zitat 5, Seite 9 in Abschnitt 1.1.1 und die Neurowissenschaftler Peter Tse (Zitat 35, Seite 63 und Robert Sapolsky (Zitat 36, Seite 63) zu verschiedenen Ansichten bezüglich der Freiheit des Willens haben, scheint es unwahrscheinlich, dass man zu einer allgemein akzeptierten Axiomatik von Theorien über den Willen gelangt. Spätestens seit David Hume ist bekannt, dass der Begriff der Kausalität mehrdeutig ist. Der von vielen Kompatibilisten verwendete Begriff des Grundes rekuriert implizit auf den der Kausalität, so dass schon allein deswegen unklar ist, warum Entscheidungen auf der Basis von Gründen freie Entscheidungen sein sollen, Habermas' Ausführungen zum vergesellschafteten Individuum oder Nida-Rümelins Begriff der humanistischen Perspektive sind überzeugend nur für den, der sowieso schon in diesem begrifflichen Rahmen denkt oder die grundsätzliche Skepsis gegenüber dem naturwissenschaftlich orientiertem Denken teilt. Wenn ich mich auf dem Weg zur Konditorei entscheide, 20 Euro an Oxfam zu spenden statt, wie ursprünglich beabsichtigt, Kuchen zu kaufen, so kann ich sagen, ich hätte das aus freien Stücken getan, niemand hätte mich gezwungen, mich so zu entscheiden. Das Gegenargument ist aber, dass ich gar nicht anders konnte, als mich so zu entscheiden, – wenn nämlich *mein* Wertesystem keine andere Entscheidung zulässt⁶⁰. Das Wertesystem ist die Ursache, der Grund⁶¹ für meine Entscheidung; man kann sagen, dass meine Handlung kausal bedingt sei. Aber deswegen ist sie noch lange nicht vorhersagbar, wie die Diskussionen über den Determinismus gezeigt haben. Es ist bemerkenswert, dass diese Diskussionen von Philosophen mit "humanistischer Perspektive" nicht berücksichtigt werden. Die Gründe, die zur meiner moralischen Entscheidung führen, müssen mir zum Zeitpunkt meiner Entscheidung wegen stochastischer Fluktuationen der Hirnaktivität nicht alle präsent sein, oder die Randbedingungen können so sein, dass die "normalen" moralischen Werte durch andere Werte bzw. Wünsche zeitlich begrenzt dominiert werden, z.B. von dem Wunsch, die Geburtstagsfeier für meinen Enkel nicht zu verderben.

⁵⁹ vergl. Wissenschaftstheorie III (2) Abschnitt 2.3

⁶⁰ Man kann höhere Werte haben, die aber bei der Erfüllung von Wünschen in einem Morast von Relativierungen verschwinden. Hier ist gemeint, dass ein Wertesystem im psychischen System wo verankert ist, dass es auch handlungsrelevant ist.

⁶¹ Ursache und Grund werden hier absichtlich als Synonyme angeführt!

1.2.7 Galen Strawsons Folgerungen

Der britische Philosoph Galen Strawson⁶² hat diese Argumentation straffer ausformuliert (Strawson, G. (2022)⁶³):

1. You do what you do – in the circumstances in which you find yourself – because of the way you then are.
2. So if you're going to be ultimately responsible for what you do, you're going to have to be ultimately responsible for the way you are — at least in certain mental respects.
3. But you can't be ultimately responsible for the way you are in any respect at all.
4. So you can't be ultimately responsible for what you do.

Strawson elaboriert insbesondere die Aussage 3., derzufolge wir letztlich nicht für die Art und Weise, in der wir sind, verantwortlich sein können. Denn wir sind nicht für unsere genetische Struktur verantwortlich, da wir diese Struktur nicht vorgeburtlich selbst bestimmt haben, und darüber hinaus sind wir nicht für unsere frühen Erfahrungen verantwortlich. Wegen unseres genetischen Erbes und der frühen Erfahrungen können wir nicht in einem späteren Lebensabschnitt eine wahre moralische Verantwortung erlangen. Denn ob man bei dem Versuch, sich selbst so zu verändern, dass man mehr moralische Verantwortung übernehmen kann Erfolg hat, hängt bereits von der Art und Weise ab, in der man wegen seines genetischen Erbes und der frühen Erfahrungen schon existiert. Weiter gibt es eine Menge von Einflußfaktoren, u.a. zufällige Einflüsse, für die man nicht verantwortlich ist:

Ultimate responsibility for how one is, is impossible, because it requires the actual completion of an infinite series of choices of principles of choice.

Strawson zitiert Nietzsche:⁶⁴

Die *causa sui*⁶⁵ ist der beste Selbst-Widerspruch, der bisher ausgedacht worden ist, eine Art logischer Notzucht und Unnatur: aber der ausschweifende Stolz des Menschen hat es dahin gebracht, sich tief und schrecklich gerade mit diesem Unsinn zu verstricken.⁶⁶

⁶²Der Sohn des renommierten britischen Philosophen Peter Frederick Strawson (1919 – 2006)

⁶³(Strawson, G. (2022) Moral responsibility without free will. *iai news*, 27th April 2022, <https://iai.tv/articles/moral-responsibility-without-free-will-auid-2115>)

⁶⁴<http://www.zeno.org/Philosophie/M/Nietzsche,+Friedrich/Jenseits+von+Gut+und+B>

⁶⁵*causa* = Grund, Ursache, *sui* = selbst, *causa sui* = Ursache seiner selbst. In der Theologie Charakterisierung Gottes, im hier gegebenen Kontext die Freiheit, etwas aus sich selbst heraus tun zu können.

⁶⁶Das Verlangen nach »Freiheit des Willens«, in jenem metaphysischen Superlativ-Verstande, wie er leider noch immer in den Köpfen der Halb-Unterrichteten herrscht, das Verlangen, die ganze und letzte Verantwortlichkeit für seine Handlungen selbst zu tragen und Gott, Welt, Vorfahren, Zufall, Gesellschaft davon zu entlasten, ist nämlich nichts Geringeres, als eben jene *causa sui* zu sein und, mit einer mehr als Münchhausenschen Verwegenheit, sich selbst aus dem Sumpf des Nichts an den Haaren ins Dasein zu ziehn; etc

Nietzsche, F. (1886): Jenseits von Gut und Böse, Erstes Hauptstück: von den Vorurteilen der Philosophen, 11 - 20.

1.2.8 Albert Einsteins Bekenntnis

Albert Einstein (im Anhang von Plancks "Where is Science going?") hat Strawsons Gedanken offenbar schon vorher gehabt:

"Honestly I cannot understand what people mean when they talk about the freedom of the human will. I have a feeling, for instance, that I will something or other; but what relation this has with freedom I cannot understand at all. I feel that I will to light my pipe and I do it; but how can I connect this up with the idea of freedom? What is behind the act of willing to light the pipe? Another act of willing? Schopenhauer once said: Der Mensch kann was er will; er kann aber nicht wollen was er will (Man can do what he wills but he cannot will what he wills)." Epilogue: A socratic dialogue Planck-Einstein-Murphy, in: Planck (1932), p. 201, Z 33

So kann man die Debatte auf den relevanten Punkt bringen, ganz ohne Perspektiven und Dualismen zu bemühen. Wahrscheinlich ist Einsteins *statement* für Philosophen wie Habermas und Nida-Rümelin eine Provokation.

1.3 Philosophie und Empirie

Auf die Frage nach dem freien Willen kann man ähnlich reagieren. Anhänger der These des freien Willens werden von Argumentationen vom Strawson-Typ nicht überzeugt, da sie die subjektive Erfahrung der Freiheit bereits als Nachweis für die Freiheit ansehen. Gerade wegen vieler Erfahrungen und der Möglichkeit rationaler Überlegungen sei es doch möglich, den Effekt genetischen Erbes und früher Erfahrungen hinter sich zu lassen, moralische Verantwortung "im eigentlichen Sinne" zu übernehmen und damit über einen freien Willen zu verfügen. Man muß aber sehen, dass derlei Aussagen nur Annahmen bzw. Hypothesen sind; empirisch arbeitende Psychologen würden beginnen, über einen Versuchsplan nachzudenken, der es erlaubt, derlei Hypothesen empirisch zu testen, und tatsächlich liefert Google Scholar auf die Anfrage "development of moral judgment" eine Fülle von Arbeiten zu diesem Thema. Philosophische Mutmaßungen über die Möglichkeiten moralischer Entwicklungen bieten allerdings wegen ihrer Abstraktheit kaum mehr als Allgemeinplätze. Darüber hinaus enthalten Arbeiten nach Art Habermas', Nida-Rümelins etc so gut wie keine Analysen, die auf die These, der Wille sei frei, führen, sie gehen im Gegenteil davon aus, dass der Wille frei ist. Begriffliche Explikationen u.a. des Konzepts 'freier Wille' sind oft hermeneutische Explorationen, die in Form apodiktischer Aussagen dokumentiert werden, man denke an Habermas' Aussage⁶⁷, in der er die Alltagspsycho-

⁶⁷ "Die objektivierende Sprache der Neurobiologie mutet dem "Gehirn" die grammatische Rolle zu, die bisher das "Ich" gespielt hat, aber sie findet damit an die Sprache der Alltagspsy-

logie zur Norm erhebt, an deren Sprache "die Neurobiologie keinen Anschluß" finde und er feststellt, dass "naturwissenschaftliche Theorien oft am Commonsense abprallen". Das tun sie gelegentlich in der Tat, aber sind Commonsense-Theorien deshalb auch wahr? Die historisch gut belegte Möglichkeit, dass der von ihm beschworene Commonsense ebensogut an wissenschaftlichen Theorien abprallen könnte, hat Habermas in seinem Aufsatz nicht in Betracht gezogen. Dass die Sonne im Osten auf- und im Westen untergeht hat zu der Commonsense-Theorie geführt, dass die Sonne sich um die Erde dreht, – sollte man deswegen Kopernikus, Galilei und andere vergessen? In der Geschichte der Philosophie des freien Willen ist "jeder Stein schon einmal gewendet" worden, wie Spohn (2008) treffend anmerkt; Habermas' Commonsense-Theorie des freien Willens ist wohl einer dieser Steine. Spohn erklärt knapp: "Jedermann bei Verstand glaubt an den freien Willen. Was immer wir genau darunter verstehen, er ist jedenfalls etwas, was wir haben." Hat Spohn recht, so war Einstein nicht bei Verstand, als er sich über den freien Willen äußerte (Zitat 33, Seite 58).

Die Definitionen von Anhängern der These des *unfreien* Willens gehen davon aus, dass Deliberationen unter anderem von nicht bewußten Aktivitäten des Gehirns beeinflusst werden können. Dieser Ansatz hat den Vorteil, mit empirischen Daten kompatibel zu sein. Auf Introspektionen beruhende Überlegungen zum freien Willen mögen als subjektiv wahr erscheinen, sind aber nicht notwendig valide (Nisbet & Wilson(1977)). Freie Entscheidungen gelten als bewußte Entscheidungen, aber wie das Gehirn es schafft, ein Ich-Bewußtsein zu erzeugen ist nicht bekannt. Noë, Dretske (vergl. Zitat 11, Seite 29), Habermas, Nida-Rümelin und andere haben ihre Aussage über die Außerhirnlichkeit des entscheidenden Ichs nicht analytisch hergeleitet, sondern als Offenbarung präsentiert, die habermassche Trennung von "Gehirn" und "Ich" soll wohl bedeuten, dass das Ich nicht vom Gehirn erzeugt wird. Behauptungen wie diese implizieren keinen Gewinn an Kenntnis oder Einsicht, sondern provozieren allenfalls die verdutzte Frage, wo sich das Ich denn sonst konstituieren soll. Das Gehirn erzeugt Gedanken, Erinnerungen, Wahrnehmungen, und Gedanken über Gedanken, gedankliche Operationen sind oft logisch, und sehr oft nicht logisch – pauschale Aussagen wie

"In dieser Dimension vollzieht sich die rationale Motivation von Überzeugungen und Handlungen nach logischen, sprachlichen und pragmatischen Regeln, die sich nicht auf Naturgesetze reduzieren lassen." (s. Habermas, p. 886)

sind glatt formulierte, aber irreführende Behauptungen: wenn sich die genannten pragmatischen Regeln "nicht auf Naturgesetze reduzieren lassen", so kann

chologie keinen Anschluss. Die Provokation, die darin besteht, dass "das Gehirn" statt meiner "selbst" denken und handeln soll, ist gewiss nur eine grammatische Tatsache; aber so schirmt sich die Lebenswelt gegen kognitive Dissonanzen erfolgreich ab. – Natürlich wäre das nicht die erste naturwissenschaftliche Theorie, die auf diese Weise am eeee Commonsense abprallt." Habermas (2004), p. 872

das einfach daran liegen, dass sie inadäquat sind, etwa weil sie sich aus einer fehlerhaften Alltagshermeneutik ergeben, – man denke an Daniel Kahnemans Buch *Thinking Fast and Slow* (2011), in dem die Systematik (die Naturgesetzlichkeit) des fehlerhaften Denkens beschrieben wird. Habermas’ Begriff des Naturgesetzes zugrunde entspricht dem Standard des 19-ten Jahrhunderts, und er verliert kein Wort zur Frage, warum sich die genannten Regeln *nicht* auf Naturgesetze reduzieren lassen.

In experimentalpsychologischen, mathematisch formulierten Modellen über das Entscheidungsverhalten, wie etwa Busemeyer & Townsends (1993) Ansatz, erlauben, die empirische Überprüfung von speziellen Hypothesen⁶⁸. Die experimentalpsychologische Forschung zu Denk- und Entscheidungsprozessen zeigt unter anderem, dass kognitive Prozesse störanfällig sind (eine allgemeine Darstellung findet man in Kahneman (2011) und Kahneman, Sibony, Sunstein (2021), Tversky & Kahneman (1974) liefern eine knappe Zusammenfassung experimenteller Befunde zu kognitiven Täuschungen insbesondere auch im Entscheidungs-, d.h. im Deliberationsverhalten, denn Entscheidungen sind ja das Resultat von Deliberationen. Diese Befunde verweisen auf die letztlich neuronale Bedingtheit der Entscheidungsfindung: was subjektiv rational (im Sinne von Habermas) erscheint, kann in der Realität äußerst irrational sein.

Das Konzept des freien Willens wird im Allgemeinen mit der Möglichkeit moralischen Handelns gekoppelt: es wird argumentiert, dass die Möglichkeit moralisch verantwortlichen Handelns die Existenz des freien Willens voraussetzt bzw. diesen impliziert. Da die (Neuro-)Biologie eine Naturwissenschaft ist und diese wiederum auf der Annahme des Kausalitätsprinzips basiere, müssen Anhänger dieser Argumentation erklären, warum der Wille frei sein kann, was insbesondere viele Kompatibilisten neue Varianten des Dualismus annehmen lässt, z.B. Perspektiven- oder Commonsensedualismen (in den Abschnitten über Nida-Rümelin und Habermas in Abschnitt 1.2.3). Im Rahmen einer *experimentellen Philosophie* sind Umfragen durchgeführt worden, in denen die intuitiven Vorstellungen der Bevölkerung über den Freien Willen und die moralische Verantwortung untersucht werden, und tatsächlich scheint kulturübergreifend eine dualistische Grundeinstellung den Glauben an einen freien Willen besser vorherzusagen als die Überzeugung, die Welt sei durchgängig deterministisch organisiert (Wisniewski et al. (2019)). Philosophen argumentieren oft, ihre philosophisch begründeten Ansichten entsprächen den Mehrheitsansichten in der

⁶⁸”Decision field theory provides for a mathematical foundation leading to a dynamic, stochastic theory of decision behavior in an uncertain environment. This theory is used to explain (a) violations of stochastic dominance, (b) violations of strong stochastic transitivity, (c) violations of independence between alternatives, (d) serial position effects on preference, (e) speed-accuracy trade-off effects in decision making, (f) the inverse relation between choice probability and decision time, (g) changes in the direction of preference under time pressure, (h) slower decision times for avoidance as compared with approach conflicts, and (i) preference reversals between choice and selling price measures of preference. The proposed theory is compared with 4 other theories of decision making under uncertainty.” Abstract zu Busemeyer & Townsend (1993)

allgemeinen Bevölkerung (Nahmias et al. (2005)). Dementsprechend wurden experimentalpsychologische Studien begrüßt, die nahelegen, dass bestätigende Aussagen über die Existenz des freien Willen und die damit einhergehende moralische Verantwortung einen positiven Effekt auf das Verhalten haben (konkret: die Aussage, es gäbe einen freien Willen, haben die Schummelrate bei kognitiven Tests reduziert, während Aussagen der Art, der Wille sei determiniert, den gegenteiligen Effekt erzeugten (Vohs et al. (2009), Baumeister et al. (2009), Baumeister et al. (2011)). Wie so oft lassen kritische Analysen nicht lange auf sich warten, etwa Nahmias (2011), Iso-Ahola (2017). Nun geht es aber bei der philosophischen Diskussion der Frage nach dem freien Willen nicht darum, ob Befürworter der These des freien Willens die moralisch höherwertigen Aufsätze schreiben, sondern um die Frage, ob sie mit ihren Argumenten recht haben. Schon bei der Betrachtung der Argumente Nida-Rümelins wurde nach dem epistemischen Nutzen philosophischer Betrachtungen gefragt, die sowohl zur der Stützung der These der Existenz eines freien Willens, wie auch zur Stützung der Gegenthese verwendet werden können.

So kommt man zu der Frage, ob die Hypothese der Existenz eines freien Willens überhaupt allein mit philosophischen Methoden entschieden werden kann. Roth (2004) hat den philosophischen, die Empirie nicht in Rechnung stellenden Ansatz, scharf kritisiert, u.a. weil die Philosophen empirische Ergebnisse zu diesem Thema entweder gar nicht erst zur Kenntnis nehmen, oder sie als irrelevant desavouieren. Insbesondere J. Habermas scheint sich von seiner schon in *Zur Logik der Sozialwissenschaften* (1970) und *Erkenntnis und Interesse* (1973) elaborierten Kritik an der Empirie und an "szientistischem" Denken nicht distanziert zu haben, und Alva Noës Glauben an den Primat der Philosophie hat die Grenze zur Bizarrie deutlich überschritten.

Es ist nicht klar, was mit dem Begriff der philosophischen Methode überhaupt gemeint ist. Man könnte an den Philosophen Hans Georg Gadamer (1900 - 2002) erinnern, der in seinem Werk "Wahrheit und Methode" die Hermeneutik als eine allgemeine Methode der Erkenntnisgewinnung vorstellte. Gadamers Ansatz ist vielleicht zu allgemein. Die Behauptung etwa Habermas', Nida-Rümelins und Noës, das Gehirn habe nichts mit der Entscheidungsfindung zu tun, folgt nicht aus einer überprüfbaren Annahme, sie schwebt frei im Raum der assoziativen Möglichkeiten. Dies gilt auch für Max Plancks (Zitat 27, Seite 49) "Punkt in der unmessbaren Welt von Geist und Materie . . . der in sich eine ganze Welt enthält und unser emotionales Leben, unseren Willen und unsere Gedanken enthält" und der nach Planck der Kausalität entzogen ist.

So hat man zwei Welten, die eine von Denkern wie G. Strawson, S. Harris, R. Sapolsky, G. Roth, W. Singer etc, und die andere von Denkern wie Kant, Nida-Rümelin, Habermas, Noë, wobei man Kant und die übrigen drei besser nicht in einer Reihe nennen sollte. Keiner der Denker der einen Welt kann einen Denker aus der anderen Welt überzeugen, weil sie, wie es scheint, von verschiedenen Interpretationen des Begriffs 'freier Wille' ausgehen. Skep-

tiker des freien Willens, d.h. Denker, die der These der Existenz eines freien Willens skeptisch gegenüber stehen, gehen anscheinend von der eingangs gegebenen, Standarddefinition von 'frei' aus, derzufolge der Wille 'unfrei' ist, wenn es Randbedingungen gibt, die letztlich die Entscheidungen bestimmen. Zu den Randbedingungen gehören auch zufällige Effekte, die in die jeweiligen Deliberationen eingehen, so dass unklar wird, was 'frei' überhaupt bedeuten soll. Die Argumente der Anhänger der These erscheinen eher als Rechtfertigung einer bereits existierenden Überzeugung.

Die Situation erinnert an das Postulat der klassischen Methodologie des rationalen Denkens⁶⁹, demzufolge nach hinreichenden Begründungen für eine Aussage gesucht werden soll, denn es gebe nur eine Wahrheit (Postulat des theoretischen Monismus). Angenommen, man möchte zeigen, dass die Aussage q : "Der Wille ist frei" wahr ist. Dazu benötigt man bereits als wahr erkannte Aussagen p , aus denen q folgt, $p \rightarrow q$. Die Aussagen p folgen wiederum aus anderen, bereits als wahr erkannten Aussagen, etc. Dann ergibt sich das von Albert so genannte Münchhausen-Trilemma: 1. Man gerät in einen infiniten Regress, oder 2. in einen logischen Zirkel in der Deduktion, d.h man greift auf Aussagen zurück, die schon begründungsbedürftig waren, oder 3. man muß den Begründungsversuch einfach abbrechen, dh man kann über einen bestimmten Punkt hinaus das Verfahren einfach nicht weiter fortführen, obwohl noch keine letztliche Begründung vorliegt. Das Trilemma ist nicht ganz unproblematisch, weil es ja ebenfalls begründet werden muß und das Trilemmaprinzip damit auf sich selbst anwendbar ist. Nun können alle hier behandelten Philosophen, einschließlich Max Plancks, sagen, ihre Argumente seien "evident". Insbesondere Noë scheint so zu denken: er präsentiert sein Extracranialargument, als sei es für jeden, der seinen Ausführungen gefolgt ist, evident wahr, die eben genannte Folge von Argumenten, die auf die Aussage 25, Seite 47, führt, würde gar nicht benötigt. Aber auch der Begriff der Evidenz ist sorgfältig der philosophischen Analyse unterzogen worden. Der Evidenzbegriff wurde von Karl R. Popper und dann insbesondere von Wolfgang Stegmüller (1954) kritisiert, der zu dem Schluß kam⁷⁰

Das Evidenzproblem ist absolut unlösbar, die Frage, ob es Einsicht
gibt oder nicht, ist absolut unentscheidbar. Z 34

In diesem Zusammenhang sei auch daran erinnert, dass Philosophen die Unmöglichkeit des *experimentum crucis* nachgewiesen haben (vergl. Wiss'theorie III(2), Abschn. 2.4). Dieses Argument ist sehr allgemein und man könnte deswegen vermuten, dass die Empirie sowieso keinen Sinn macht, – eine Sichtweise, die vor allem immer wieder bei Habermas angedeutet wird, für den die philosophische Reflexion die zentrale Methode der Annäherung an die Frage nach dem freien Willen ist. Natürlich kann man der Philosophie ein grundsätz-

⁶⁹Wissenschaftstheorie III(2), Abschnitt 2.3

⁷⁰Wissenschaftstheorie III (2), Abschnitt 2.2

liches Primat zuschreiben, denn die Entscheidung, sich bestimmten Fragen nicht nur theoretisch, sondern empirisch zuzuwenden, sei letztlich eine philosophische Entscheidung, weil sie letztlich auf dem philosophisch gewonnenen Argument beruht, mit rein philosophischen Mitteln sei die Welt nun einmal nicht zu ergründen. Aber diese Diskussion muß an dieser Stelle nicht wiederholt werden, sie ist in Wissenschaftstheorie II und III eingiermaßen ausführlich geführt worden. Die Fortschritte der Wissenschaften hängen bekanntlich nicht nur von im Lehnstuhl erdachten und dann "schön" formulierten Theorien ab, sondern wesentlich auch von kluger Empirie, so dass die Aussage über die Möglichkeit von Entscheidungsexperimenten *cum grano salis* betrachtet werden muß. Schon in der Einleitung Abschnitt 1.1 wurde angemerkt, dass es müßig sein könnte, einen Streit um die Existenz oder Nicht-Existenz des freien Willens zu führen, einfach, weil zu jeder Aussage, die zur Begründung der These der Existenz eines freien Willens gemacht wird, eine Gegenthese formuliert werden kann, der zufolge es keinen freien Willen gibt.

Entgegengesetzte Ansichten in den Naturwissenschaften: Die, was den freien Willen angeht, gegensätzlichen Auffassungen der Physiker G. t'Hooft und A. Zeilinger sind bereits in Abschnitt 1.1.1 zitiert worden. Gegensätzliche Aussagen gleicher Art findet man aber auch bei Neurowissenschaftlern:

Peter Tse (1962 –): "Just as Watson and Crick showed that there is a physical basis, a physical mechanism—namely DNA—that could account for evolution," Tse says, "I am arguing that rapid synaptic reweighting is the physical mechanism that gives humans the power to exercise free will."
Tse (2013), Z 35

Der ebenfalls renommierte Neurowissenschaftler Robert Sapolsky (1957 –) kommt in Sapolsky (2023) zu dem entgegengesetzten Schluß:

"The world is really screwed up and made much, much more unfair by the fact that we reward people and punish people for things they have no control over," Sapolsky said. "We've got no free will. Stop attributing stuff to us that isn't there." (2023)⁷¹, Z 36

So stellt sich die Frage, ob es eine empirische Möglichkeit gibt, die Hypothese der Existenz eines freien Willens zu überprüfen. Dazu ergeben sich die folgenden Bemerkungen.

Verifikation, Falsifikation, und der freie Wille In Wissenschaftstheorie III (2) ist die Möglichkeit, Hypothesen entweder zu verifizieren oder zu falsifizieren bereits diskutiert worden. Bei den folgenden Betrachtungen geht es darum, zu zeigen, dass die Hypothese des freien Willens grundsätzlich weder verifiziert noch falsifiziert werden kann. Schon Kant hat geschrieben, dass der Begriff der Freiheit kein empirischer Begriff sein könne. Im Folgenden geht es um den freien Willen an sich, also nicht um den freien Willen als Konsequenz noch als

⁷¹<https://phys.org/news/2023-10-scientist-decades-dont-free.html>

Bedingung für die Fähigkeit zu moralischem Handeln. Offenbar ist die Möglichkeit einer Verifikation bzw. Falsifikation an gewisse Vorbedingungen geknüpft, nämlich dass irgendwelche Indikatoren I für eine Entscheidung E existieren, – oder nicht existieren. Man hat also das Schema

- *Verifikation*: Zeige, dass dem Indikator I von E *keine* kausalen Effekte K_I vorausgehen,
- *Falsifikation*: Zeige, dass I kausal erzeugt wird, d.h. dass kausalen Effekte K_I existieren.

Der Schluß, dass die Hypothese H_{fw} des freien Willens nicht testbar ist, ergibt sich, wenn gezeigt werden kann, dass weder die Verifikation – der Nachweis der Existenz kausaler Effekte – noch die Falsifikation – der Nachweis der Nichtexistenz kausaler Effekte – möglich ist. Grundsätzliches zur Problematik der Verifikation findet man schon bei David Hume und sehr ausführlich bei Karl R. Popper, und ebenso Grundsätzliches zur Problematik der Falsifikation ist bereits in der Kritik an Poppers Philosophie der Falsifikation formuliert worden. Hier soll noch auf weitere Aspekte der Verifikation-Falsifikation-Problematik hingewiesen werden.

Northcott (2018) argumentiert zu Recht, dass der Nachweis der Nichtexistenz von Prozessen K_I kaum möglich ist, denn ein Nachweis der Nichtexistenz von Prozessen K_I bedeutet ja, dass das Zustandekommen von E nicht erklärbar ist, und Nichterklärbarkeit für sich genommen beweise noch nichts, insbesondere beweise sie nicht, dass irgendwelche K_I -Prozesse nicht in der Zukunft gefunden werden können, – es sei denn, dass man *beweisen* kann, dass K_I -Prozesse nicht existieren können. Northcott zitiert dazu den britischen Physiker J.S. Bell (1964), der eine Ungleichung – die nach ihm benannte *Bellsche Ungleichung* – aufgestellt hat, die erfüllt sein muß, wenn Variablen existieren, die bestimmte Phänomene kausal erklären. Diese Variablen müssen nicht explizit bestimmbar sein, weshalb von *versteckten* Variablen (hidden variables) die Rede ist. In der klassischen Physik wird die Ungleichung stets erfüllt, nicht aber in der Quantenphysik. Die Ungleichung wird beim Phänomen der Quantenverschränkung verletzt, was als Nachweis der Nichtexistenz von versteckten, Kausalität reflektierenden Variablen gedeutet wird. Für die Nichtexistenz von K_I -Prozessen liegt aber kein derartiger Beweis vor, was als Nichtverifizierbarkeit von H_{fw} gedeutet werden kann. Es liegen auch keine Daten vor, die einen empirischen Hinweis auf Nichtkausalität liefern, wie man sie für die Widerlegung der Behauptung, die MMR-Impfungen würden Autismus erzeugen, hat (MMR: Masern, Mumps und Röteln)⁷². Northcott verweist auf die Vernetztheit der Hirnprozesse, die den Nachweis der *Nichtexistenz* von Prozessen, die kausal zu Entscheidungen führen, schwierig, wenn nicht unmöglich machen.

⁷²Die methodischen Fehler, die zu dieser Behauptung führten, werden in <https://www.chop.edu/centers-programs/vaccine-education-center/vaccines-and-other-conditions/vaccines-autism> erläutert.

Um H_{fw} zu falsifizieren muß man nachweisen, dass man stets Prozesse finden kann, die eine Indikatoraktivität I bewirken. Als ein solcher Nachweis wird oft das Bereitschaftspotential B_p genannt, das der Entscheidung E , die Hand zu heben, vorangeht. Nun wird aber B_p nicht in jedem Versuchsdurchgang beobachtet, d.h. B_p ist ein probabilistisches Phänomen. Schurger et al. (2016) haben dementsprechend argumentiert, dass ein B_p eher eine Art von Hintergrundrauschen (background noise) repräsentiere. Streng genommen müsste eine inkompatibilistische Position eingenommen werden: H_{fw} gilt als widerlegt, wenn für alles Geschehen der Determinismus gilt, d.h. wenn nichts ohne Ursache geschieht. Aber hier greift die Argumentation van Kampens (Kap. IV, Abschnitt 2.4, p. 44), derzufolge ein solcher Determinismus weder bewiesen noch widerlegt werden kann. Deshalb kann man vermuten, dass H_{fw} prinzipiell nicht falsifiziert werden kann.

Die hier vorgetragene Argumentation zur H_{fw} ist sicherlich eher eine Skizze, Northcott hat sie weiter elaboriert, worauf hier nicht weiter eingegangen werden muß; es sei aber noch einmal auf Wissenschaftstheorie III (2), Abschnitt 2.4, verwiesen, wo die generelle Unterbestimmtheit empirischer Befunde (Duhem-Quine-These) diskutiert wird.

So ist es interessanter, zu untersuchen, *wie* Menschen jeweils zu ihren Entscheidungen gelangen. Die Behauptung einiger Philosophen, dass die (Natur-)Wissenschaften Fragen wie die nach dem freien Willen gar nicht beantworten können, weshalb ein nicht-szientistischer bzw. ein humanistischer Ansatz die geeignete Perspektive sei, um die Möglichkeiten eines freien Willens zu erforschen, unterliegt ebenfalls der oben erwähnten Trilemma-Problematik, die den Absolutheitsanspruch philosophischer Ansätze relativiert. Man könnte dem philosophischen Argument der Nichtwiderlegbarkeit auch entgegenhalten, es sei eine *petitio principii*, also eine "Vorwegnahme des Beweisgrundes"⁷³: der Begriff des freien Willens wird so definiert, dass er als evident wahr, aber nicht weiter beweis- oder widerlegbar erscheint.

Weitere biologische Befunde Ein beliebtes philosophisches Argument zur Stützung von H_{fw} beschwört die Emergenz kognitiver Aktivitäten auf der Basis von neuronaler Aktivität, die Bewußtsein und die Freiheit des Willens ermögliche. Allerdings führt der Hinweis auf Emergenz kaum zu neuen Einsichten, so lange man sich keine expliziten Vorstellungen davon macht, welcher Art die biologischen Prozesse sind, die zur Emergenz etwa von Bewußtsein führen. In diesem Zusammenhang sei noch einmal an die Arbeit von Wegner (2003) erinnert. Die von ihm zitierten Arbeiten legen nahe, dass das bewußte Erlebnis, eine Handlung herbei geführt zu haben, eine Täuschung ist (s. die Bemerkungen zum Reduktionismus bei Habermas). Es gibt weitere Ansätze, Prozessen der Emergenz auf die Spur zu kommen, etwa die *Dynamic Core* und *Global Workspace* Hypothesen: Edelman, Gally & Baars (2011) liefern einen Überblick.

⁷³so die Übersetzung ihm Rahmen der Logik: Kodakow, N.I.: Wörterbuch der Logik, Leipzig 1983, p. 373

Die genannten Theorien basieren auf dem Ansatz des *Neural Darwinism* (Edelman (1978), 1987), Edelman & Tononi (2007)). Diesen Autoren zufolge besteht das Gehirn aus sehr vielen Gruppen von interagierenden Subpopulationen von Neuronen, deren Struktur – gemeint ist die Vernetzung der Neuronen untereinander – durch Wechselwirkungen mit der Umwelt zum Teil verstärkt, zum Teil geschwächt werden. Wesentlich für diese Entwicklung ist das Re-entry-Prinzip, demzufolge es einen stetigen, reziproken Signalaustausch zwischen den Gruppen gibt. Allgemein wird angenommen, dass das Bewußtsein u.a. das Resultat derartiger *reentranter neuronaler Aktivität* ist; dies ist insbesondere die zentrale These der *Dynamic Core-Hypothese* (Baars (1988)). Für spezifische bewußte Zustände bzw. Ereignisse sind anscheinend Aktivitäten des Thalamus und des cerebralen Kortex zuständig; Edelman, Gally, Baars (p. 4) führen den Begriff der *behavioral trinity* ein: (1) Biologische Mechanismen und phänomenale Erfahrung, (2) eine Basis für das "subjektive Selbst", (3) die Beziehung zwischen den Signalen (Inputs) dem Gehirn und dem Körper und dem eigenen Inneren (Gedächtnis), – die verschiedenen Systeme des Gehirns müssen im *dynamischen Kern* (Dynamic Core) koordiniert werden. Im dynamischen Kern werden Inputs aus der sensorischen Wahrnehmung und der aus der eigenen Bewegung mit Gedächtnisinhalten kombiniert und durch Aktivitäten aus subkortikalen Strukturen (Cerebellum, Basalganglien) moduliert.

Alle diese Befunde und Interpretationen von Befunden beruhen im Wesentlichen auf den Ergebnissen von fMRI-Untersuchungen, wobei nicht verschwiegen werden soll, dass es Kritik insbesondere an früheren fMRI-Untersuchungen gibt, bei denen Korrelationen zwischen der Aktivität in bestimmten Voxeln⁷⁴ und kognitiver Aktivität berechnet wurden; diese können Methodenartefakte sein (Vul et al. (2009)). Alternativ dazu wurde versucht, Voxel zu bestimmen, bei denen die neuronale Aktivität für gegebene kognitive Aktivitäten besonders deutlich sind, wobei über räumliche Bereiche gemittelt wird. Bei diesem Vorgehen wird aber die *differentielle Aktivierung* in verschiedenen Voxeln nicht berücksichtigt. Es hat sich jedoch gezeigt, dass Kognitionen auf der Basis von Aktivitäten in Mustern von Voxeln klassifiziert werden können. Einen Überblick liefern Norman et al (2006), oder Mahmoudi et al. (2012), die anders als Norman et al. auch eine Einführung in die neueren Methoden der Klassifikation von Mustern wie etwa Support-vector-Machines präsentieren (eine detailliertere Einführung findet man in Bishop (2006)).

Eine wichtige Frage bezieht sich auf das Erleben von Qualia und dem Selbst. Edelman et al. argumentieren, dass sie sich aus Unterscheidungen ergeben: Hitze ist nicht "grün", "grün" ist nicht "Berührung", etc. Das Erleben einer komplexen Szene ergibt sich demnach aus Unterscheidungen im Dynamischen Kern. Dabei ergibt sich das von Chalmers (1996) so genannte "schwierige Problem" (the hard problem), wissenschaftlich zu erklären, worin bewußtes Erleben be-

⁷⁴Ein Voxel ist ein 3-dimensionales *Volumenelement* (volume element) im Gehirn, im Unterschied zum *Pixel*, das ein 2-dimensionales Bildelemente (picture element) ist.

steht bzw. wie es sich entwickelt. Philosophen und Forscher, die sich die chalmersche Sichtweise zu eigen gemacht haben, sind der Auffassung, dass es sich grundsätzlich der wissenschaftlichen Forschung entzieht.

Grundsätzliche Aussagen dieser Art sind allerdings problematisch, da nicht klar ist, aus welchen als "wahr" erkannten Erfahrungen bzw. Annahmen sie folgen. Lange galt es als wahr, dass sich Leben nicht aus der bis dahin bekannten Physik und Chemie erklären lasse. Zur Erklärung des Lebens wurde eine spezielle *vis vitae*, eine Lebenskraft, postuliert. Wie so oft läßt sich diese Lehre bis in die Antike zurückverfolgen; sie hat sich bis ans Ende des 19-ten und den Beginn des 20-ten Jahrhunderts erhalten. Erst dann hatten sich Physik und Chemie so weit entwickelt, dass klar wurde, dass eine solche spezielle Kraft nicht postuliert werden muß. Chalmers hat nicht *bewiesen*, dass das Bewußtsein nicht biologisch erklärt werden kann, er hat diese These nur suggeriert, seine Argumente sind eher intuitiver Art. Man kann sich fragen, warum zum Beispiel Chalmers' Zombie-Argument zwingend sein soll⁷⁵, und dieselbe Frage stellt sich für Habermas' und Nida-Rümelins Thesen zur Extracranialität des Geistes.

Nach der Theorie des Neuronalen Darwinismus und des Dynamischen Kerns ergeben sich die Qualia aus Interaktionen höherer Ordnung; die Unterscheidung von 'warm' und 'blau' ist nur möglich, wenn die notwendigen neuronalen Strukturen vorhanden sind. Für Philosophen wie Habermas, Nida-Rümelin und vor allem Alva Noë muß allerdings gar nicht auf die Frage nach deren Existenz und Struktur eingegangen werden, weil ihrer Meinung nach das Gehirn für alles Psychische keine Rolle spielt. Die Hirnforschung findet aber Strukturen, die für diese Wahrnehmungen und Unterscheidungen in Frage kommen und im Rahmen des *Third-Person-View*-Ansatzes untersucht werden können, d.h. es ist möglich, die Korrelation zwischen subjektiver Erfahrung und kausalen neuronalen Prozessen zu entschlüsseln; Edelman et al. sehen darin die Auflösung der Barriere des chalmersschen *hard problems*. Um die Beziehung zwischen Hirnaktivität und Bewußtsein zu entschlüsseln muß aber nicht bei den Komplexitäten des menschlichen Bewußtseins beginnen. Es kann sinnvoll sein, zunächst die Beziehungen zwischen sensorischem Input und korrespondierender Wahrnehmung untersuchen, wie es etwa Cabanac et al. (2009) getan haben. Diese Autoren gehen davon aus, dass die Dimension *angenehm – unangenehm* eine Dimension des Bewußtseins von mehreren sein könnte. Das Verhalten des Organismus wird so angelegt, dass sein Zustand möglichst 'angenehm' ist. Insgesamt gehen Cabanac et al. von vier Dimensionen aus. Die erste Dimension wird als qualitativ angenommen: die Farbe rot, oder blau, oder was der Farbkreis sonst noch hergibt. Ein Geschmack ist süß, oder sauer, etc. Die zweite Dimension ist quantitativ,

⁷⁵Ein Zombie ist nach Chalmers eine molekulargenaue Kopie eines Menschen, die allerdings ohne das Bewußtsein, ohne phänomenale Erfahrung auskommt. Mit der begrifflichen Konstruktion eines solchen Wesens will Chalmers zeigen, dass das Bewußtsein keine physikalische Basis haben kann (!). Er begründet diese Argumentation mit der Aussage, Zombies seien konzipierbar (conceivable), – ergo metaphysisch möglich (Chalmers (1997)). Die philosophische Debatte darüber hält an.

sie beschreibt die Intensität eines Stimulus: die Farbe ist hell, der Ton ist laut, und so weiter. Die dritte Dimension ist affektiv ("hedonisch": angenehm – unangenehm). Die vierte Dimension ist die Dauer, sie repräsentiert die Zeit, die ein Stimulus einwirkt. Man kann dann zuerst die Anatomie betrachten. Die Generierung von Bewußtsein erfordert neuronales Gewebe, also sollte die Größe des Gehirns eine Rolle spielen, – sie wird durch den *Enzephalisationsquotienten* (EQ) angegeben. Hier gibt es einen deutlichen Sprung von Ektothermen zu Endothermen Organismen. Die Endothermen sind warmblütige, tachymetabolische⁷⁶ Tiere mit einem zehnmal so großen Gehirn als die kaltblütigen bradymetabolischen⁷⁷. Wirbeltiere mit demselben Body-Mass-Index. Ectothermische Tiere wie Lissanphibia (= rezente Amphibien wie Blindwühlen, Schwanz- und Froschlurche) haben 5 Vesikeln am cephalischen Ende eines neuronalen Tubus; bei den Lissanphibia behält das Telencephalon die Struktur eines embryonischen Vesikels bei, während man bei den Lepidosauria – eine Teilklasse von Reptilen, die zuerst im frühen Perm (Beginn ca vor 298 Millionen Jahren bis ca. vor 252 Millionen Jahren) erscheinen) – findet man einen größeren Wechsel zu einer neuen Struktur, nämlich einen Kortex, d.h. zwischen Fröschen und Eidechsen gibt es einen qualitativen Sprung, der mit einem dazu korrespondierenden Sprung in der Produktion des Neurotransmitters Dopamin einhergeht, der wiederum mit einer hedonistischen Erfahrung gekoppelt ist, und hedonistische Erfahrungen sind bewußte Prozesse. Wie Cabanac ausführt geht dieser Sprung mit Veränderungen im Verhalten einher. Weitere Details können hier nicht im Detail besprochen werden, aber zusammenfassend kann man sagen, dass es bei diesen Veränderungen nicht um das Bewußtsein eines Selbst geht, sondern nur um die Existenz eines mentalen Raumes (Cabanac et al. (2009), p. 271, Abbildung 6), wobei es sich um einen "quantitativen" (Cabanac et al.), nicht "qualitativen" Prozess handelt, dh dieses erste Bewußtsein entwickelt sich in einem inkrementellen Prozess.

Mashour et al. (2013) haben ebenfalls einen Überblick über Arbeiten zur Frage, wann Bewußtsein entstanden ist, geliefert. Sie untersuchen u.a. die Möglichkeit, die Beziehung zwischen Bewußtsein und Anästhesie zu untersuchen um Einblick in die Phylogenie des Bewußtseins zu erlangen und bestätigen bisherige Befunde, denen zufolge die Entstehung zumindest eines repräsentierenden Bewußtseins phylogenetisch weit zurückgreift. Die grundsätzlichen Mechanismen des menschlichen Bewußtseins sind phylogenetisch sehr früh in der Geschichte der Wirbeltiere entstanden. Die Unterschiede zwischen den Spezies bezüglich des Bewußtseins sind gradueller und nicht qualitativer Natur. Der Beginn der Entwicklung liegt wahrscheinlich in der Kopplung eines internen Systems, das die Befriedigung von Notwendigkeiten wie Nahrung etc regelt, und einem situationsbezogenen Bewußtseinszustand (gemeint ist das englische 'awareness') besteht. Es zeigt sich, dass der Ort der Entstehung von Bewußtsein in den phylogenetisch

⁷⁶tachymetabolisch = mit erhöhtem Grundumsatz

⁷⁷bradymetabolisch = hoher aktiver Stoffwechsel und ein erheblich langsamerer Ruhestoffwechsel

alten Hirnteilen wie dem Hirnstamm, dem Thalamus und dem Hypothalamus etc liegt, während der Neokortex relativ wenig am "primitiven" Teil des Bewußtseins teilhat. Bei "höherem" Bewußtsein zeigen PET-Untersuchungen zusätzlich eine verteilte Aktivität im gesamten Hirn auf; Mashour et al. liefern eine Menge von Details. Ob man aus immer genauer werdenden Untersuchungen diejenigen Details der Mechanismen herausanalysieren kann, die zumindest 'einfaches' Bewußtsein erzeugen, ist gegenwärtig noch nicht klar. Die epistemische Situation scheint vergleichbar zu sein mit der in der Physik, als man noch keine Idee hatte, was Gravitation eigentlich ausmacht, wenn auch die newtonsche Formel ziemlich genau beschreibt, wie sie wirkt. Dann aber dann kam Einstein und erklärte, dass Gravitation ein Effekt der Krümmung des Raums ist, die durch die Verteilung von Masse im Raum verursacht wird. Das Entstehen von Bewußtsein läßt sich vielleicht durch molekuarbiologische Prozesse erklären, oder – man kann es nicht voraussagen – durch eine Neufassung der quantenmechanischen Erklärung von Roger Penrose, und Singers Aussage "Unser Gehirn setzt seine eigenen kognitiven Werkzeuge ein, um sich selbst zu begreifen, und wir wissen nicht, ob dieser Versuch gelingen kann" (Zitat 12, Seite 29) gilt nach wie vor. Aber Patricia Smith Churchland hat bereits gemahnt, sich nicht in rein philosophischen Betrachtungen zu verlieren:

The philosophical lesson is this: when not much is known about the topic, don't take terribly seriously someone else's heartfelt conviction about what problems are scientifically tractable. Learn the science, do the science, and see what happens.

Patricia Smith Churchland, *Brainwise* (2002), p. 180

2 Anhang

2.1 Turingmaschinen

Die meisten Menschen beantworten die Frage, wieviel 13 mal 17 ist, nicht aus dem Gedächtnis, sondern sie rechnen, meistens im Kopf, etwa $10 \cdot 13 = 130$, plus $7 \cdot 13 = 70 + 21 = 91$, also $130 + 91 = 221$. Oder $17 \cdot 10 = 170$, plus $3 \cdot 17 = 30 + 21 = 51$ und $170 + 51 = 221$. Das Wichtige dabei ist, dass jedes Mal eine Folge von Operationen durchgeführt wird, Zwischenergebnisse abgespeichert werden, etc. Diese Folgen können sich von Person zu Person ein wenig unterscheiden, eine gegebene Person wird vermutlich stets ein und dieselbe Folge wählen, aber sie kann auch zwischen Folgen wählen. bei größeren Zahlen wird sie mit Papier und Bleistift arbeiten und dabei die Operationen in stets der gleichen Reihenfolge vornehmen. Folgen dieser Art sind Algorithmen. Turing (1936) hat diese Art, Aufgaben zu lösen, als Vorbild für eine allgemeine Art, Algorithmen zu definieren, gewählt. Der Begriff des Algorithmus geht auf den iranisch-stämmigen Mathematiker und Universalgelehrten al-Chwarizmi zurück, der um das Jahr 800 herum im Haus der Weisheit in Bagdad wirkte. Die Definition eines Algorithmus genügte aber nicht den Anforderungen an die Genauigkeit in der modernen Mathematik, so dass der englische Mathematiker Alan Turing ((1912 – 1954) eine präzise, formale Definition vorlegte. Dazu konzipierte er eine, heute nach ihm bekannte Maschine – eben die Turing-Maschine. Sie besteht zunächst aus einem in Felder eingeteilten, unendlichen (Papier-)Band. In jedes Feld kann ein Zeichen aus einem Arbeitsalphabet geschrieben werden. Über dem Band befindet sich ein Schreib- und Lesekopf, der jeweils ein Zeichen liest, oder liest, oder schreibt, wobei die Aktivität des Kopfes durch ein Program gesteuert wird. Der Kopf kann seine Position um ein Feld nach links oder nach rechts verändern. Die Definition der Turing-Maschine ist dann

Definition 1 Eine Turing-Maschine (TM) ist durch das 7-Tupel

$$M = (Z, \Sigma, \Gamma, \delta, z_0, \square, E)$$

- gegeben, wobei Z eine endliche Zustandsmenge ist,
2. Σ das Eingabealphabet,
3. $\Gamma \supset \Sigma$ das Arbeitsalphabet,
4. $\delta : Z \times \Gamma \longrightarrow Z \times \Gamma \times L, R, NP'$ die Überfunktionsfunktion,
5. $z_0 \in Z$ der Startzustand,
6. $\square \in \Gamma - \Sigma$ ein Leerzeichen, 7. $E \supseteq Z$ ist die Menge der Endzustände,

Dann bedeutet

$$\delta(z, a) = (z', b, x)$$

δ beschreibt die Funktionsweise der Turing-Maschine. Befindet sich M im Zustand z und unter dem Lesekopf das Zeichen a steht, so geht M im folgenden Schritt in den Zustand z' über, schreibt das Zeichen b auf die Position, auf der

vorher a stand und führt dann eine Bewegung des Lesekopfes $x \in \{L, R, N\}$ aus L = Bewegung nach links, R = Bewegung nach rechts, N = neutral, d.h. der Kopf wird nicht bewegt. Die TM ist eine allgemeine Beschreibung dessen, was ein Mensch macht, wenn er eine Aufgabe einer Folge von Regeln löst, und sie ist deswegen ein generisches Modell für einen Computer. Und sie ist, wie am Anfang dieses Abschnitts schon erwähnt, eine allgemeine Beschreibung von Prozessen.

s. Schöning, p. 72.

Algorithmen und der Begriff der Turing-Maschine, eher intuitiv, der Begriff der Berechnung.

2.2 Turings Halte-Problem

Man stelle sich vor, dass man ein Programm P für irgendetwas geschrieben hat. Man habe dabei irgendwo einen Fehler gemacht, der bewirkt, dass das Programm in eine Schleife ("loop") läuft und deswegen ewig weiterläuft, sofern man es nicht abbricht. Man weiß aber nicht, ob es nur so lange läuft, weil es in einer Schleife läuft, oder ob die Berechnung, die das Programm ausführen soll, so langwierig ist. So entsteht der Wunsch nach einem Programm P_0 , das prüft, ob P in eine Schleife laufen wird, oder nicht. In diesem Zusammenhang ist es wichtig, einen weiteren Begriff einzuführen, den der *universellen Turing-Maschine*. Dies ist eine Turing-Maschine, die eine beliebige Turing-Maschine für einen beliebigen Input simuliert.

Nun ist P_0 selbst ein Programm, das einen Fehler enthalten kann, der impliziert, dass P_0 in eine Schleife laufen kann. Man könnte auf die Idee kommen, dass ein weiteres Programm P_1 geschrieben werden muß, das prüft, ob P_0 in eine Schleife läuft. Dieser Ansatz ist aber keine Lösung, weil sofort klar ist dass dann ein Programm P_2 benötigt wird, das prüft, ob P_1 in eine Schleife läuft, etc, d.h. man gerät in eine Art von infinitem Regress. Dann aber bleibt nur die Möglichkeit, P_0 mit sich selbst zu prüfen, d.h. P_0 zum Input von P_0 zu machen. P_0 muß sich selbst prüfen, – womit man sich aber die Falle der Selbstreferenz einhandelt.

Ein Programm ist die Implementierung eines Algorithmus, d.h. eines Verfahrens, das ein Problem in einer endlichen Folge von Schritten löst. Das Verfahren kann deterministisch oder probabilistisch sein. Bei einem deterministischen Algorithmus liegt nach jedem Schritt eindeutig fest, welche Operation als nächste durchgeführt wird, bei einem probabilistischen Algorithmus wird die jeweils folgende Operation mit einer bestimmten Wahrscheinlichkeit aus einer Menge von möglichen Operationen gewählt.

Definition Turingmaschine und universelle Turingmaschine hier?

Turing (1936), Abschnitt 8., hat gezeigt, dass ein Programm P_0 , das bestimmt, ob es in eine Schleife gerät oder nicht existiert. Diese Aussage ist

das Halte-Theorem.

Turings Beweis ist indirekt, dh es wird angenommen, dass ein solches Programm P_0 (Turing (1939) hat dafür den Ausdruck 'Orakel' eingeführt) existiert, um dann zu zeigen, dass diese Annahme auf einen Widerspruch führt, d.h. das Orakel existiert nicht. Eine grundlegende Annahme für die folgenden Betrachtungen ist, dass die Arbeit eines Programms stets aus einer diskreten Folge von Operationen oder Schritten besteht; es kann gezeigt werden, dass diese Annahme keine Einschränkung darstellt, d.h. das Halte-Theorem gilt auch für kontinuierliche Prozesse (s. unten).

Definitionsgemäß liefert das Orakel P_0 auf die Frage, ob ein gegebenes Programm P_s für einen gegebenen Input P (der selbst ein Programm sein kann) hält (H für 'hält') oder in eine Schleife gerät (L für 'Schleife' (loop)) stets die korrekte Antwort. Die hier interessierenden, möglichen Antworten von einem Programm (einschließlich P_0 und P_s) sind also

$$P = \begin{cases} L, & P \text{ gerät in eine Schleife,} \\ H, & P \text{ hält nach endlich vielen Schritten} \end{cases}$$

Die Frage von P_s an das Orakel P_0 ist, ob P , auf sich selbst angewandt, hält oder nicht. Es soll dann die folgende Regel gelten:

Regel R: Wenn das Orakel sagt, dass P loopt, stoppt P_s . Wenn das Orakel sagt, P wird stoppen, loopt P_s .

Nun kommt die Frage: was geschieht, wenn P_s sich selbst als Input bekommt? Die möglichen Antworten sind

1. Das Orakel sagt, P_s wird für immer loopen. Nach der Regel R muß P_s dann stoppen.
2. Das Orakel sagt, P_s wird stoppen. Nach der Regel R muß P_s dann loopen.

Resultat: Das Orakel gibt in jedem Fall die falsche Antwort, entgegen der Annahme, dass es stets die Wahrheit spricht. Damit ist die Annahme, dass das Orakel stets die Wahrheit spricht widerlegt. In anderen Worten: das Orakel existiert nicht, - es gibt kein Programm, das voraussagen kann, ob ein Programm einen Fehler enthält oder nicht, d.h. es existiert kein Programm, das sich selbst "erkennt"-

Man kann diesen Sachverhalt anhand von Cantors Diagonalverfahren illustrieren. Dazu werde die folgende Notation eingeführt: $P_1(P)$ heißt, dass P_1 das Programm P prüft. Für $P = P_1$ erhält man $P_1(P_1)$: P_1 prüft sich selbst.

Man hat nun zwei Möglichkeiten: $P_1(P_1)$ stoppt (Stop), d.h. $P_1(P_1) = H$, oder läuft in eine Schleife (Loop), $P_1(P_1) = L$. Die Menge der möglichen Programme ist abzählbar unendlich (da es aus abzählbar vielen Elementen besteht). Jedes Programm kann als Input für jedes Programm gewählt werden und stoppt dann entweder (H für hält), oder läuft in eine Schleife (L für loop). Man kann dann eine Tabelle anschreiben: Es sei noch einmal betont, dass die Tabelle *alle* möglichen Programme mit allen möglichen Inputs enthalten soll. Die i -te

	P_1	P_2	P_3	P_4	P_5	$\dots$
P_1	L	H	H	L	L	$\dots$
P_2	H	L	H	H	L	$\dots$
P_3	H	L	H	L	H	$\dots$
P_4	H	L	H	L	L	$\dots$
P_5	H	L	H	H	H	$\dots$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\ddots$
Diagonale	L	L	H	L	H	
inv. Diag	H	H	L	H	L	

Zeile, etwa die für P_3 enthält die "Reaktionen" des dritten Programms auf alle möglichen Inputs: für P_1 stoppt es (H), für P_2 loopt es (L) etc. Die Tabelle scheint also alle möglichen Reaktionen, – wie es scheint. Nun betrachte man die Reaktionen in den Diagonalzellen der Tabelle: $L, L, H, L, H, \dots$ Dies sind die Reaktionen der Programme auf sich selbst. Nun ersetze man jedes Symbol durch das alternative Symbol, also L durch H und H durch L . Man erhält die Folge $H, H, L, H, L; \dots$ Diese Folge kann nicht in der Tabelle vorkommen, denn sie unterscheidet sich von den Folgen in der Tabelle in Bezug auf mindestens ein Element, nämlich das Element in der Diagonalzelle. Dies bedeutet, dass es unmöglich ist, ein Programm zu schreiben, für das sich diese neue Folge ergibt. Dieser Sachverhalt spricht gegen die Existenz des Orakel-Programms, – Weder P_5 noch das Orakel können existieren! Damit ist das Halte-Theorem bewiesen.

2.3 Lloyds Beweis

First: check notion of Universal Turing machine! Es sei T eine monoton wachsende Abbildung der natürlichen Zahlen in die natürlichen Zahlen, also $T : \mathbb{N} \rightarrow \mathbb{N}$, $\mathbb{N}$ die Menge der natürlichen Zahlen. Weiter seien $|d|$ und $|k|$ die Längen (in bits) der Zahlen d und k . $d_T(k)$ bezeichne einen zeitbegrenzten Entscheider (d von decider), mit

$$d_T(k) = \begin{cases} d(k), & d \text{ liefert Output auf Input } k \\ & \text{in } T(|d| + |k|) \text{ oder weniger Schritten} \\ 0, & \text{sonst} \end{cases}$$

Keine Antwort in $T(|d| + |k|)$ bedeutet 'Nein'. Wendet man das Diagonalargument auf die Frage an, ob zeitbegrenzte Entscheider mit 'ja' oder 'nein' in weniger als T Zeiteinheiten entscheiden, so zeigt sich, dass sie im Allgemeinen *mehr* als T Zeiteinheiten benötigen. Zeit wird dabei in Termen von komputationalen Schritten repräsentiert.

Es werden nun zeitbegrenzte Turingmaschinen $\mathcal{T}_d^T$ mit dem Input k und dem Output $d_T(k)$ betrachtet. Es sei $f(d_T, k) = d_T(k)$; f ist die Antwort auf die Frage, welche Entscheidung durch einen typischen Entscheider innerhalb

der Zeit T gegeben wird. Es läßt sich nun eine Turingmaschine konzipieren, die (d, k) als Input akzeptiert und die $f(d_T, k)$ als Output liefert; f kann als Antwort auf die Frage, wie ein typischer Entscheider in der Zeit T handelt, gesehen werden. Die Funktion f ist, wenn nicht explizit auf die Zeit T Bezug genommen wird, *berechenbar* in dem Sinne, dass sie durch eine Turingmaschine mit Input (d, k) und Output $f(d_T, k)$ berechnet werden kann. Sie ist aber *nicht* in der Zeit T berechenbar, d.h. jede Berechnung für die Entscheidung, die ein Entscheider treffen wird, nimmt mehr Zeit in Anspruch als die Entscheidung des betrachteten Entscheiders.

Um zu zeigen, dass die Berechnung von f mehr als T Einheiten erfordert, werde die Matrix A_T betrachtet, mit dem (d, k) -ten Element $f(d_T, k)$; dies ist die Matrix aller Entscheidungsfunktionen $d_T(k)$, die in der Zeit T berechnet werden können. Es sei $g(k) = 0$ wenn $f(k, k) = 1$, und umgekehrt. Dann folgt $g(k) = \neg f(k, k)$ ($\neg$ = nicht). Ist f in der Zeit T berechenbar, so ist es auch g . Wenn aber g in der Zeit T berechenbar ist, dann ist notwendig $g(g) = f(g, g)$. Damit hat man aber einen Widerspruch, da ja $g(g) = \neg f(g, g)$. Folglich sind weder f noch g in der Zeit T berechenbar.

Betrachte einen Entscheider d (d von *decider*). Versehe d mit einem Input: die Beschreibung eines anderen Entscheiders d' . d' entspricht einer Turingmaschine T_d mit zwei Input-Bändern, einem für d' und das andere für die Spezifikation des Entscheidungsproblems k . $d(d', k)$ sei der Output von T_d , wenn T_d stoppt, und $d(d', k) = d(k)$, wenn $d' = 0$, und $d(d', k) = d'(k)$ sonst. Wenn T_d nicht hält, ist der Output undefiniert. Das heißt, der universelle Entscheider d kann entweder die "straight decision" $d(k)$ machen, oder die Entscheidung eines anderen Entscheiders d' (einschließlich von sich selbst, $d = d'$, machen.

Anwendung auf die Selbstbeobachtung eines Entscheiders;

Unentscheidbarkeit ergibt sich, wenn der universelle Entscheider Aussagen über seine eigene Zukunft machen soll. Es werde eine 3-dimensionale Matrix mit den Elementen $d(d', k)$ betrachtet, wobei

$$d(d', k) = \begin{cases} d(d', k), & d(d', k) \text{ ist definiert,} \\ F, & d(d', k) \text{ ist nicht definiert} \end{cases}$$

Für festes k werden nun die Diagonalelemente $d' = d$ betrachtet. Dies sind die Reaktionen des Entscheiders d auf sich selbst. Hier gilt das Diagonalisierungsargument für das Halte-Problem (s. Anhang). Es impliziert, dass die Funktion $f_k(d) = d(d', k)$ gilt, und $f_k(d) = F$ sonst, und $f_k(d)$ ist nicht berechenbar. Das heißt, der Entscheider d kann keine Aussage machen, wenn es um seine eigene, zukünftige Entscheidung geht.

Mit T_d^T werden nun zeitabhängige Turingmaschinen bezeichnet. Diese liefern Outputs $d_T(d', k)$, wobei cccc

$$d_T(d', k) = \begin{cases} d(d', k), & T_d^T \text{ hält nach Zeit } T(|d| + |d'| + |k|) \\ 0, & \text{sonst} \end{cases}$$

Nun betrachte man die 3-dimensionale Matrix mit den Elementen $d_T(d'k)$. Für festes k betrachte man weiter die Diagonalelemente mit $d = d'$. Aus dem Hartmanis-Stearns Diagonalverfahren folgt, dass für diesen Fall die Outputs nicht in der Zeit T berechnet werden können. Dies bedeutet, dass ein universaler Entscheider, der Aussagen über seine eigenen Entscheidungen machen soll, intrinsisch weniger Effizient ist als ein Entscheider, der Entscheidungen ohne die Introspektion treffen soll.

,

Literatur

- [1] Albert, H. (2009) Die dualistische Metaphysik von Jürgen Habermas – Eine kritische Untersuchung seines nachmetaphysischen Denkens’. RMM Vol. 0, Perspectives in Moral Science, ed. by M. Baurmann & B. Lahno, 2009, 109-120
- [2] An der Heiden U., Chaos und Ordnung, Zufall und Notwendigkeit, in: Küppers G. (Hrsg.), Chaos und Ordnung - Formen der Selbstorganisation in Natur und Gesellschaft, Reclam, Stuttgart 1996
- [3] Baars, B.J. A cognitive theory of consciousness. Cambridge University Press 1988
- [4] Bak P, Tang C, Wiesenfeld K (1988) Self-organized criticality. *Phys Rev A* 38, 364–374
- [5] Bandura, A., & Walters, R.H.: Social learning and personality development. Holt Rinehart and Winston: New York 1963
- [6] Baumeister, R.F., E.J. Masicampo, and C.N. DeWall, (2009) Prosocial benefits of feeling free: Disbelief in free will increases aggression and reduces helpfulness. *Personality and Social Psychology Bulletin*, 35, 260–268
- [7] Baumeister, R. F., Crescioni, A. W., & Alquist, J. L. (2011). Free will as advanced action control for human social life and culture. *Neuroethics*, 4, 1-11.
- [8] Bell, J. S. (1964) On the Einstein Podolsky Rosen Paradox. In: *Physics*, Band 1, Nr. 3, 1964, S. 195–200
- [9] Bishop, C.M.: Pattern Recognition and Machine Learning. Springer Verlag 2006
- [10] Brans, C.H. (1988) Bell’s Theorem Does Not Eliminate Fully Causal Hidden Variables. *International Journal of Theoretical Physics*, 27 (2), 219–226
- [11] Braun, M. N., Wessler, J., & Frieze, M. (2021). A meta-analysis of Libet-style experiments. *Neuroscience & Biobehavioral Reviews*, 128, 182-198.
- [12] Braun, H. A. (2021). Stochasticity versus determinacy in neurobiology: From ion channels to the question of the "free will". *Frontiers in Systems Neuroscience*, 15, 629436.
- [13] Busemeyer, J. R., & Townsend, J. T. (1993) Decision field theory: a dynamic-cognitive approach to decision making in an uncertain environment. *Psychological Review*, 100(3), 432-59

- [14] Cabanac, M., Cabanac A.J., Parent, A. (2011) The emergence of consciousness in phylogeny. *Behavioral Brain Research*, 198, 267 – 272
- [15] Carnap, R.: Scheinprobleme in der Philosophie und andere metaphysikkritische Schriften. Philosophische Bibliothek 560 (1928)
- [16] Chalmers, D. J. (1997). The conscious mind: In search of a fundamental theory. Oxford Paperbacks.
- [17] Clark, A., Chalmers. D.J. (1998), The extended mind. *Analysis*, 58 (1): 7–19.
- [18] Conway, J., & Kochen, S. (2006). The free will theorem. *Foundations of Physics*, 36(10), 1441-1473.
- [19] Desmurget, M., Reilly, K. T., Richard, N., Szathmari, A., Mottolese, C., & Sirigu, A. (2009). Movement intention after parietal cortex stimulation in humans. *Science*, 324(5928), 811-813.
- [20] Dretske, F. (1989). Reasons and causes. *Philosophical Perspectives*, 3, 1-15
- [21] Earman, J. (1986). A primer on determinism (Vol. 37). Springer Science & Business Media.
- [22] Edelman, G. M., Gally, J. A., & Baars, B. J. (2011). Biology of consciousness. *Frontiers in psychology*, 2, 8906
- [23] Epilogue: A socratic dialogue Planck–Einstein–Murphy, in: Planck (1932)
- [24] Ellis, G. (2020). From chaos to free will. *Aeon Magazine*, 12. <https://aeon.co/essays/heres-why-so-many-physicists-are-wrong-about-free-will>
- [25] Frankfurt H.G. (1971). Freedom of the Will and the Concept of a Person. *The Journal of Philosophy*, Vol. 68, No. 1 (Jan. 14, 1971), pp. 5-20
- [26] S. J. Freedman, J. F. Clauser (1972) Experimental Test of Local Hidden-Variable Theories. *Physical Review Letters*, 28 (14) 938–941
- [27] Fried, I., Mukamel, R., & Kreiman, G. (2011). Internally generated preactivation of single neurons in human medial frontal cortex predicts volition. *Neuron*, 69(3), 548-562.
- [28] Goldstein, S., Tausk, D. V., Tumulka, R., & Zanghì, N. (2010). What does the free will theorem actually prove. *Notices of the AMS*, 57(11), 1451-1453.
- [29] Granicky, M.S. (1995) Universal computation and other capabilities and continuous dynamical systems. *Theoretical Computer Science*,

- [30] Haggard, P., & Eimer, M. (1999). On the relation between brain potentials and the awareness of voluntary movements. *Experimental brain research*, 126, 128-133. 138, 67-100
- [31] Griffel, F. A.: Gesellschaft und Verantwortung: Zu Vorträgen von Max Planck über Willensfreiheit. *Archiv für Rechts- und Sozialphilosophie / Archives for Philosophy of Law and Social Philosophy*, 1998, Vol. 84, No. 4 (1998), 517-528
- [32] Habermas, J. (2004) Freiheit und Determinismus, *Deutsche Zeitschrift für Philosophie*, Berlin 52, 6, 871-890
- [33] Hartmanis, J., & Stearns, R. E. (1965). On the computational complexity of algorithms. *Transactions of the American Mathematical Society*, 117, 285-306.
- [34] Haggard, P., & Eimer, M. (1999). On the relation between brain potentials and the awareness of voluntary movements. *Experimental brain research*, 126, 128-133
- [35] Haggard, P. (2008). Human volition: towards a neuroscience of will. *Nature Reviews Neuroscience*, 9(12), 934-946.
- [36] Haggard, P. (2009). The sources of human volition. *Science*, 324(5928), 731-733.
- [37] Harris, S. Free Will, Free Press, New York 2012
- [38] He, B. J. (2014). Scale-free brain activity: past, present, and future. *Trends in cognitive sciences*, 18(9), 480-487.
- [39] Heisenberg, W. (1927). Über den anschaulichen Inhalt der quantentheoretischen Kinematik und Mechanik. *Zeitschrift für Physik*, 43(3-4), 172-198.
- [40] Heisenberg, W.: Physikalische Prinzipien der Quantentheorie (1958), 2. unveränd. Nachdr., Berlin Spektrum, Akad. Verlag 2001
- [41] Heisenberg, M. (1983) Initiale Aktivität und Willkürverhalten bei Tieren, *Naturwissenschaften*, 70, 70 – 78
- [42] Heisenberg, M. (2009) Is free will an illusion? *Nature*, 459(7244), 164–165.
- [43] Horsthemke, W., Levever, R. (1984). Noise induced transitions. Theory and applications in physics, Chemistry, and biology. Springer-Verlag, Berlin, Heidelberg, New York, Tokyo 1984
- [44] Hossenfelder, S., Palmer, T. (2020). Rethinking superdeterminism. *Frontiers in Physics*, 8, 1 – 24

- [45] Hossenfelder, S.: Existential Physics – A Scientist’s Guide to Life’s Biggest Questions. Viking, New York 2022
- [46] Hume, D.: ”Of liberty and necessity”, in Book 2 von *A Treatise of Human Nature* (1739) sowie in *An Enquiry concerning Human Understanding* (1748)
- [47] Hume, D. (1789). Of probability; and of the idea of cause and effect (L. A. Selby-Bigge, Ed.). In D. Hume & L. A. Selby-Bigge (Ed.), *A treatise of human nature* (Reprinted from the original ed., pp. 73–78). The Clarendon Press
- [48] Iso-Ahola, S. E. (2017). Reproducibility in psychological science: When do psychological phenomena exist?, *Frontiers in Psychology*, 8, 237147.
- [49] James, W. (1884). The dilemma of determinism (pp. 114-140). Kessinger Publishing.
- [50] Johnson-Laird, P. N. (1997). Rules and illusions: A critical study of Rips’s The Psychology of Proof. *Minds and Machines*, 7(3), 387-407.
- [51] Kahneman, D. Thinking fast and slow, New York 2021
- [52] Kahneman, D., Sibony, O., Sunstein, C.R.: Noise – a flaw in human judgment. Dublin 2021
- [53] Kampen van, N. G. (1991) Determinism and Predictability, *Synthese*, 89 (2) p.273
- [54] Kobeleva, X., López-González, A., Kringelbach, M. L., & Deco, G. (2021). Revealing the relevant spatiotemporal scale underlying Whole-Brain dynamics. *Frontiers in neuroscience*, 15, 715 – 861
- [55] Kornhuber, H. H., & Deecke, L. (1965) Hirnpotentialänderungen bei Willkürbewegungen und passiven Bewegungen des Menschen: Bereitschaftspotential und reafferente Potentiale. *Pflüger’s Archiv für die gesamte Physiologie des Menschen und der Tiere*, 284, 1-17
- [56] Larsson, Jan-Åke (2014), Loopholes in Bell inequality tests of local realism. *Journal of Physics A: Mathematical and Theoretical*, 47(42): 16. arXiv:1407.0363
- [57] Libet, B. Mind Time. The temporal factor in consciousness. Harvard University Press 2004
- [58] Libet B, Gleason CA, Wright EW, Pearl DK (1983) Time of conscious intention to act in relation to onset of cerebral activity (readiness-potential): the unconscious initiation of a freely voluntary act. *Brain*, 106, 623–642, 1985

- [59] Libet, B.: Unconscious cerebral initiative and the role of conscious will in voluntary action. *The Behavioral and Brain Sciences*, 9 8, 1985, S. 529–566.
- [60] Libet, B.: Mind Time. The temporal factor in consciousness. Harvard University Press 2004
- [61] Lloyd, S., A Turing test for free will. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 2012, 370(1971), 3597–3610.
- [62] Lucas, T. (1961) Minds, Machines and Gödel. *Philosophy*, 36 (137) 112–127
- [63] Mahmoudi, A., Takerkart, S., Regragui, F., Boussaoud, D., & Brovelli, A. (2012). Multivoxel pattern analysis for fMRI data: a review. *Computational and mathematical methods in medicine*, 2012(1), 961257.
- [64] Marley, A. A. J., & Colonius, H. (1992) The "horse race" random utility model for choice probabilities and reaction times, and its comparing risks interpretation. *Journal of Mathematical Psychology*, 36(1), 1–208
- [65] Mashour, G. A., & Alkire, M. T. (2013). Evolution of consciousness: phylogeny, ontogeny, and emergence from general anesthesia. *Proceedings of the National Academy of Sciences*, 110 (supplement₂), 10357 – 10364
- [66] Moral responsibility without free will. *iai news*, 27th April 2022, <https://iai.tv/articles/moral-responsibility-without-free-will-auid-2115>
- [67] Murray, J.D.: Mathematicial Biology, Springer-Verlag, Berlin, Heidelberg etc 1989
- [68] Nahmias, E., Morris, S., Nadelhoffer, T., Turner J. (2005) Surveying Freedom: Folk Intuitions about Free Will and Moral Responsibility. *Philosophical Psychology*, (18(5), 561 — 584
- [69] Nahmias, E. (2011) Why 'Willusionism' Leads to 'Bad Results': Comments on Baumeister, Crescioni, and Alquist. *Neuroethics*, 4, 17–24
- [70] Nida-Rümelin, J., Freiheit und Kausalität. Redigierte Tonband-Mitschrift der frei gehaltenen Akademievorlesung am 12. April 2007 an der Berlin-Brandenburgischen Akademie der Wissenschaften. Redaktion: Christine Bratu, M.A., Universität München.
- [71] Nietzsche, F.: www.zeno.org/Philosophie/M/Nietzsche,+Friedrich/Jenseits+von+Gut+und+B
- [72] Nietzsche, F. (1886): Jenseits von Gut und Böse, Erstes Hauptstück: von den Vorurteilen der Philosophen. München 1954
- [73] Nisbet, R.E., Wilson, T.D. (1977) Telling more than we can now: verbal reports on mental processes. *Psychological Review*, 84, 231 — 259

- [74] Noë, A.: Du bist nicht dein Gehirn – Eine radikale Philosophie des Bewußtseins. München 2010, Originaltitel *Out of our heads*, New York 2009
- [75] Norman, K. A., Polyn, S. M., Detre, G. J., & Haxby, J. V. (2006). Beyond mind-reading: multi-voxel pattern analysis of fMRI data. *Trends in cognitive sciences*, 10(9), 424-430.
- [76] Northcott, R. (2019). Free will is not a testable hypothesis. *Erkenntnis*, 84(3), 617-631
- [77] Penrose, R.: The Emperor's New Mind. Oxford University Press, New York 1989
- [78] Penrose, R.: Shadows of the Mind. London 1995
- [79] Planck, M.: Where is Science going?, New York 1932
- [80] Planck, M.: Scheinprobleme der Wissenschaft. Von Max Planck – Vortrag gehalten in Göttingen am 17. Juni 1946
- [81] Rachlin, H. (2012). Is the mind in the brain? A review of: Out of our heads: Why you are not your brain, and other lessons from the biology of consciousness by Alva Noë (2009). *Journal of the Experimental Analysis of Behavior*, 98(1), 131.
- [82] Rauch, D., Handsteiner, J., Hochrainer, A., Gallicchio, J., Friedman, A. S., Leung, C., ... & Zeilinger, A. (2018). Cosmic Bell test using random measurement settings from high-redshift quasars. *Physical review letters*, 121(8), 080403
- [83] Roth, G. (2014), Willensfreiheit, Physik und Hirnforschung. *Studium generale: Auf dem Weg zu einem allgemeinen Teil der Wissenschaften*, 163-174.
- [84] Roth, G. (2005), Gehirn, Gründe, Ursachen. *DZPhil* 53, 5, 691-705
- [85] Rowlands, M., Lau, J., Deutsch, M.: "Externalism About the Mind", The Stanford Encyclopedia of Philosophy (Winter 2020 Edition), Edward N. Zalta (ed.),
URL = <https://plato.stanford.edu/archives/win2020/entries/content-externalism/>;
- [86] Rubinov, M., Sporns, O., Breakspear, J.P.M. (2011) Neurobiologically Realistic Determinants of Self-Organized Criticality in Networks of Spiking Neurons *PLOS, Computational Biology*, 7(6), 1–14
- [87] Russell, P., "Hume on Free Will", The Stanford Encyclopedia of Philosophy (Fall 2017 Edition), Edward N. Zalta (ed.),
URL = <https://plato.stanford.edu/archives/fall2017/entries/hume-freewill>

- [88] Sapolsky, R. Determined. The Science of Life without Free Will. Vintage Penguin-Random House 2023
- [89] Sapolsky, R. <https://phys.org/news/2023-10-scientist-decades-dont-free.html>
- [90] Scheibe, E.: Die Philosophie der Physiker. München 2006
- [91] Schopenhauer, A.: Die beiden Grundprobleme der Ethik, Preisschrift über die Freiheit des Willens (1838)
- [92] Schöning, U. : Theoretische Informatik – kurz gefasst. Spektrum Akademischer Verlag, Heidelberg 2009
- [93] Schurger, A., Mylopoulos, M., & Rosenthal, D. (2016). Neural antecedents of spontaneous voluntary movement: A new perspective. *Trends in Cognitive Sciences*, 20(2), 77–79
- [94] Singer, W. (2004) Selbsterfahrung und neurobiologische Fremdbeschreibung. *Deutsche Zeitschrift für Philosophie*, 52, 2, 235-256.
- [95] Soon, C. S., Brass, M., Heinze, H. J. & Haynes, J. D. (2008). Unconscious determinants of free decisions in the human brain. *Nature neuroscience*, 11(5), 543-545.
- [96] Spohn, W. (2008) Der Kern der Willensfreiheit. In: Sturm, D. (Hg.) Julian Nida-Rümelin über Vernunft und Freiheit. de Gruyter 2008
- [97] Srinivasan, R., Russell, D. P., Edelman, G. M., & Tononi, G. Increased synchronization of neuromagnetic responses during conscious perception. *Journal of Neuroscience*, 1999, 19(13), 5435-5448.
- [98] Srinivasan, R., Winter, W. R., Ding, J., & Nunez, P. L. EEG and MEG coherence: measures of functional connectivity at distinct spatial scales of neocortical dynamics. *Journal of neuroscience methods*, (2007), 166(1), 41-52.
- [99] Strawson, G. (2022) Moral responsibility without free will. iai news, 27th April 2022, <https://iai.tv/articles/moral-responsibility-without-free-will-auid-2115>
- [100] R.L Stevenson – Life in Western Samoa bzw In the South Seas
- [101] Tetens. H.: Kants Kritik der Reinen Vernunft - Ein systematischer Kommentar. Reclams Universal Bibliothek, Stuttgart 2006
- [102] t'Hooft, G. T. (2017), Free will in the theory of everything. arXiv preprint arXiv:1709.02874.

- [103] 't Hooft, G. (2020). Deterministic quantum mechanics: the mathematical equations. *Frontiers in Physics*, 8, 253.
- [104] Tse, P.U.: The Neural Basis of Free Will: Criterial Causation. MIT-Press 2013. Auch:
<https://pbs.dartmouth.edu/news/2013/03/neuroscientist-says-humans-are-wired-free-will>
- [105] Turing, A. M. (1936). On computable numbers, with an application to the Entscheidungsproblem. *J. of Math*, 58(345-363), 5.
- [106] Turing, A. M. (1939). Systems of logic based on ordinals. Proceedings of the London Mathematical Society, Series 2, 45, 161-228.
- [107] Tversky, A., & Kahneman, D. (1974). Judgment under Uncertainty: Heuristics and Biases: Biases in judgments reveal some heuristics of thinking under uncertainty. *science*, 185(4157), 1124-1131
- [108] Van Kampen, N. G. (1991). Determinism and predictability. *Synthese*, 89, 273-281.
- [109] Vicente, J. G. (2023). Emergence of universal computations through neural manifold dynamics. bioRxiv, 2023-02.
- [110] von Weizsäcker, C.F.: Erinnerungen an Martin Heidegger. In: Fetzer, G. (Hrsg.) Das Lesebuch der Deutschen. München 1988, 182 – 189
- [111] Varela, F. J., Thompson, Rosch, E.: The Embodied Mind: Cognitive Science and Human Experience. MIT Press 1991
- [112] Vohs, K.D. and J.W. Schooler. 2008. The value of believing in free will: Encouraging a belief in determinism increases cheating. *Psychological Science*, 19: 49–54
- [113] Vul et al. (2009)⁷⁸
- [114] von Weizsäcker, C.F., Boone, C.C. (1977) Werner Heisenberg. *CrossCurrents*, Winter 1977-8, 27 (4), 417-425
- [115] von Weizsäcker, C.F.: Große Physiker – Von Aristoteles bis Werner Heisenberg. München 1999
- [116] Vorberg, D., & Ulrich, R. (1987). Random search with unequal search rates: Serial and parallel generalizations of McGill's model. *Journal of Mathematical Psychology*, 31(1), 1-23

⁷⁸Vul, E., Harris, C., Winkielman, P., & Pashler, H. (2009). Voodoo correlations in social neuroscience. *Perspectives on psychological Science*, 4(3), 274-290.

- [117] Walter, W. G., Cooper, R., Aldridge, V. J., McCallum, W. C., & Winter, A. L. (1964). Contingent negative variation: an electric sign of sensorimotor association and expectancy in the human brain. *Nature*, 203(4943), 380-384.
- [118] Wegner D.M. (2003) The mind's best trick: how we experience conscious will. *Trends in cognitive Sciences*, 7(2w), 65 – 69
- [119] Wisniewski D, Deutschländer R, Haynes, J-D (2019) Free will beliefs are better predicted by dualism than determinism beliefs across different cultures. PLoS ONE 14(9): e0221617. <https://doi.org/10.1371/journal.pone.0221617> (open source)
- [120] Zeilinger, A. The Dance of the photons. 2010

Index

- Algorithmus, 70
- Antinomien
 - der reinen Vernunft, 21
- Apperzeption, 23
- Argumentationen, rekursive, 52
- Aussagen
 - apriorische, 20
- Autonomie
 - vernünftige, 21, 22
- behavioral trinity, 66
- Bellsche Ungleichung, 64
- Bereitschaftspotential, 15
 - lateralisiertes, 16
- Commonsense-Dualismus, 41
- Deliberation, 4
- Dualismus
 - epistemischer, 34
 - Perspektiven-, 34
- Dualismus
 - Common sense, 41
- Dynamic Core-Hypothese, 66
- dynamischen Kern, 66
- Dämon
 - Laplace'scher, 49
- Einstein, Albert, 8
- Ellis, George F.E., 11
- Emergenz, 65
- Enzephalisationsquotient, 68
- EQ, 68
- Erklärungslücke, 46
- experimentum crucis, 62
- Gerard t'Hooft, 9
- Heisenberg, Werner, 7
- Idealismus
 - transzendentaler, 24
- intelligible
 - Dinge an sich, 23
- initiale Aktivität, 26
- Inkompatibilisten, 6
- Kausalität aus Freiheit, 22
- Kompatibilisten, 5
- Lehrsätze, vernünftelnde, 21
- Letztbegründungsproblem, 56
- Libertarianer, 5
- Libets Experimente, 15
- Monismus
 - theoretischer, 62
- Neural Darwinism, 66
- neuronale Aktivität
 - reentrante, 66
- Nietzsche, 57
- Noumena, 5
- noumenale Welt, 23
- Noumenon, 23
- Perspektive
 - humanistische, 41
- petitio principii, 65
- Philosophie
 - experimentelle, 60
- Psychophysik, 44
- Quantenfluktuationen, 10
- Quantenverschränkung, 64
- Raum der Gründe, 28
- Raum der Ursachen, 28
- Reduktionismus, 37
- Scheinproblem
 - nach M. Planck, 50
- Spontaneität
 - Idee einer, 22
- Sprachspiel, 35
- Stevenson, 14
- subjektives Selbst, 66
- Superdeterminismus, 8, 9

szientistisch/nicht-szientistisch, 32

Taboo-Bruch, 14

Third Person View, 67

Turing-Maschine, universelle, 71

Umwelt

 soziolinguistische, 47

Urteil

 synthetisches, 20

Urteile

 syntthetisch a priori, 24

van Kampens Theorem, 10

Vergesellschaftung, 37

Vernunft

 Kant, 20

vernünftelnLe Lehrsätze, 21

Verstand

 Kant, 20

vis vitae, 67

Wille

 höherer Ordnung, 4

Willkürfreiheit, 41

Zustandsübergänge

 rauschinduzierte, 11

Zwei-Welten-Lehre, 23

Zwei-Welten-Theorie, 5