

DIE AUTONOMIE DES MENSCHEN UND DER MASCHINE

G E G E N W Ä R T I G E
D E F I N I T I O N E N _ V O N
A U T O N O M I E _ Z W I S C H E N
P H I L O S O P H I S C H E M
H I N T E R G R U N D _ U N D
T E C H N O L O G I S C H E R
U M S E T Z B A R K E I T

ERDUANA SHALA

**DIE AUTONOMIE DES MENSCHEN UND DER MASCHINE –
GEGENWÄRTIGE DEFINITIONEN VON AUTONOMIE ZWISCHEN
PHILOSOPHISCHEM HINTERGRUND UND TECHNOLOGISCHER UMSETZBARKEIT**

KARLSRUHE 2014

Hochschulschrift. Eingereicht 2012 am KIT als Masterarbeit, überarbeitet und online veröffentlicht 2014.

Vorwort

Der vorliegende Text, im Jahr 2012 als wissenschaftliche Arbeit zur Erlangung des akademischen Grades einer Magistra Artium am Institut für Philosophie des Karlsruher Instituts für Technologie entstanden, endet mit der Feststellung:

Es ist der Mensch, der sowohl seine eigene Autonomie, als auch die der Maschine erfunden hat. Es liegt daher in seiner Verantwortung stets für ein klares Verständnis zu sorgen, was im Laufe des technischen Fortschritts unter diesem Begriff aufgefasst werden soll.

Diese Forderung sollte sich kaum ein Jahr später im Auftrag eines BMBF-Projekts widerspiegeln: 2013 wurde das Fraunhofer Institut für System und Innovationsforschung ISI durch das BMBF mit der Durchführung des Projekts *Wandel von Autonomie und Kontrolle in Mensch-Technik-Interaktionen* beauftragt, um jene begrifflichen Wandlungstendenzen auszuloten. Die Ergebnisse dieses Projekts, an dem ich maßgeblich mitgewirkt habe, sind bereits publiziert und auch als Online-Heuristik verfügbar. (Siehe <http://www.amtir-heuristik.de/> sowie Gransche / Shala / Hubig et al.: *Wandel von Autonomie und Kontrolle durch neue Mensch-Technik-Interaktionen*, Karlsruhe 2014)

Aufgrund der weiterhin bestehenden Wichtigkeit der Thematik und einiger Nachfragen nach der Masterarbeit, möchte ich diese in der vorliegenden Form der Öffentlichkeit zugänglich machen. Diese redigierte Version unterscheidet sich sehr geringfügig vom ursprünglichen Text.

Großer Dank gilt an dieser Stelle Prof. Dr. Dr. Mathias Gutmann sowie Prof. Dr. Armin Grunwald für die Betreuung der Arbeit, deren Inhalte weiterhin Früchte tragen sowie für die Zustimmung zur Veröffentlichung.

Erduana Shala

Inhaltsverzeichnis

Siglenverzeichnis.....	IV
1. Einleitung	1
2. Der Autonomiebegriff in der Philosophie	5
2.1. Der Autonomiebegriff.....	5
2.2. Der Autonomiebegriff in der Antike.....	5
2.3. Aristotles‘ Ethik.....	6
2.4. Autonomie als Forderung der Philosophie der Aufklärung	10
2.4.1. Autonomie bei Kant	11
2.3.2. Reaktionen auf Kants Autonomieverständnis	15
2.5. Ausdifferenzierung des kantischen Autonomiebegriffs bis in die heutige Zeit	17
2.6. Zusammenführung und Fortgang.....	20
3. Autonomie der Technologien.....	23
3.1. Begriff und Gegenstand	23
3.1.1. Das Attribut ‚autonom‘	23
3.1.2. Gegenstand	25
3.2 Ebenen der Mechanisierung und Autonomiegrade	27
3.3. Kriterien technischer Autonomie	29
3.3.1. Kontrolle.....	29
3.3.2. Intelligenz.....	31
3.3.2.1. Herkunft des Intelligenzbegriffes.....	31
3.3.2.2. Definition Intelligenz heute.....	33
3.3.2.3. Künstliche Intelligenz	34
(1) „Acting Humanly: The Turing Test approach”	34
(2) „Thinking humanly: The cognitive modeling approach”	36
(3) „Thinking rationally: The ‘laws of thought’ approach”	36
(4) „Acting rationally: The rational agent approach“.....	37
3.4. Zusammenführung und Fortgang.....	38
4. Autonomie kommt nur dem Menschen zu	41
4.1. Sinnhaftigkeit von autonomen Handlungen.....	41

4.1.1. Persönlichkeit und Moralität als Leitmotiv für autonomes Handeln	42
4.1.2. Zur Pragmatik der Autonomie des Handelns	46
4.2. Mentale Phänomene, Qualia	48
4.3. Bewertung und Zusammenführung	52
5. Erweiterung der Autonomie des Menschen	54
5.1. Aufgabenzuteilung und Zuschreibung von Autonomie	54
5.1.1. Missverständliche Metaphorik?	55
5.1.2. Das „Autonomie-Paradoxon“	58
5.2. Autonomisierung der Technologien im Kontext von Technikgenese und humaner Selbstkonstitution	61
5.2.1. Neue Mittel, alte Zwecke	62
5.2.2. Neue Mittel, neue Zwecke?	65
6. Schlussbemerkung	70
Literaturverzeichnis	VI

Siglenverzeichnis

Primärliteratur

- NE** – Aristoteles: Nikomachische Ethik, auf der Grundlage der Übersetzung von Eugen Rolfes, hrsg. von Günther Bien, 4. durchges. Aufl., Hamburg 1985
- KrV** – Immanuel Kant: Kritik der reinen Vernunft, nach der 1. und 2. Originalausgabe, hrsg. von Jens Timmermann, Hamburg 1998
- KpV** – Immanuel Kant: Kritik der praktischen Vernunft, mit einer Einleitung, Sachanmerkungen und einer Bibliographie von Heiner F. Klemme, hrsg. von Horst D. Brandt und Heiner F. Klemme, Hamburg 2003
- KU** – Immanuel Kant: Kritik der Urteilskraft, mit Einleitungen und Bibliographie, hrsg. von Heiner F. Klemme, Hamburg 2009
- MdS** – Immanuel Kant: Metaphysik der Sitten, unveränderter Abdruck 1966 der 4. Auflage 1922, hrsg. von Karl Vorländer, Hamburg 1966

Sekundärliteratur

- HWPph** – Historisches Wörterbuch der Philosophie, unter Mitw. von mehr als 1200 Fachgelehrten in Verbindung mit Günther Bien, völlig Neubearb. Ausg. d. <Wörterbuchs der philosophischen Begriffe> von Rudolf Eisler, hrsg. von Joachim Ritter, Bd. 1-13, Basel 1971-2007
- MEL** – Meyers Enzyklopädisches Lexikon, 25 Bde., 9. Aufl., Mannheim 1971-1979

1. Einleitung

Werden Maschinen¹ mit Begriffen beschrieben, die ursprünglich der Beschreibung des Menschen dienen, so liegt zunächst der Verdacht nahe, dass jene Maschinen spezifisch-menschliche Fähigkeiten oder Eigenschaften besitzen. Für körperliche Fähigkeiten, die mechanisch nachgeahmt werden, hat sich in der Alltagssprache eine anthropomorphisierende Sprechweise bereits etabliert. So wird kaum in Frage gestellt, dass bestimmte Maschinen weben, backen, sich bewegen oder arbeiten können. Bei nicht-körperlichen Eigenschaften, etwa kognitiver, sozialer oder moralischer Art sieht dies jedoch anders aus. Dass mittlerweile intelligente und rechnende Maschinen im alltäglichen Sprachgebrauch Eingang gefunden haben, wäre jedoch undenkbar ohne den langjährigen Diskurs über Künstliche Intelligenz, welcher insbesondere die zweite Hälfte des vergangenen Jahrhunderts geprägt hat.² In jüngster Zeit ist es der Autonomiebegriff, welcher zunehmend Verwendung zur Beschreibung neuer Technologien findet, wie etwa „autonome mobile Roboter“ oder „autonome Systeme“.

Dem Begriff nach rekurriert die „Autonomie“ jener Technologien auf eine bestimmte Art technologischen Fortschritts, die von der Fähigkeit zur Selbstgesetzgebung herrührt. Dies wirft aus philosophischer Sicht jedoch die Frage auf, wie die Selbstgesetzgebung in diesem Fall definiert ist, zumal sich der Autonomiebegriff in der Philosophie auf die politische oder moralische Selbstgesetzgebung von Menschen oder Menschengruppen beziehungsweise ihre Handlungen bezieht. Im *Handbuch Robotik*³ hingegen führt der Autor geradezu beiläufig die Bezeichnung „autonom“ ein, indem er prognostiziert, dass „[...] autonome Roboter in Zukunft sogar einen Großteil der Altenbetreuung übernehmen werden.“⁴

Hierin zeichnen sich bereits die vielfältigen Probleme ab, auf welche ich im Verlauf dieser Arbeit eingehen werde. Es müsste geklärt werden, inwieweit die Autonomie des Roboters mit der Ausführung der Aufgabe zusammenhängt und ob der Roboter in derselben Weise autonom ist wie der Mensch. Wird hierbei lediglich induktiv von rein

¹ Maschine (gr. *mechané* für Hilfsmittel, Werkzeug, Kriegsmaschine; lat. *machina*; frz. *machine* für Kriegs- u. Belagerungsmaschine) bezeichnet eine Vorrichtung, mit der eine Energieform für einen bestimmten Zweck in eine andere umgewandelt wird, oder „mit der die von einer Kraft-M. gelieferte Energie in gewünschte Arbeit umgesetzt wird.“ Siehe MEL, Artikel Maschine

² Dennoch ist bei diesen Vorgängen, die auf dem Vorbild mentaler Fähigkeiten des Menschen fußen, immer noch Interpretationsspielraum, ob es sich um Nachahmung oder Nachbildung von Fähigkeiten handelt. Hierfür hat sich John Searles begriffliche Unterscheidung in „weak AI“ und „strong AI“ etabliert. Siehe John R. Searle: *Minds, brains, and programs*, in: *The Behavioral and Brain Sciences*, 3 (1980), 417-457

³ Matthias Haun: *Handbuch Robotik*, Berlin / Heidelberg 2007

⁴ Haun 2007, 7

äußerlich ersichtlichen Verhaltensweisen auf eine tatsächliche Fähigkeit geschlossen oder wird der Maschine mehr Leistungsfähigkeit unterstellt als sie erbringen kann?

Da die Vernetzung und gegenseitige Abhängigkeit von Mensch und Technik zunimmt, stellt sich daher insbesondere die Frage, welche Auswirkungen die zunehmende Verflechtung von Mensch und Maschine auf unser Verständnis von Autonomie hat. Leisten autonome Technologien einen Beitrag zur Definition der Autonomie des Menschen oder handelt es sich um voneinander unabhängige Begriffe?

Erst eine differenzierte Betrachtung der Autonomie des Menschen erlaubt es, diverse Definitionsmöglichkeiten einer Autonomie der Maschine oder auch einen erweiterten Autonomiebegriff für den Menschen plausibilisieren zu können. Daher soll zunächst in Kapitel 2 untersucht werden, wie sich der Autonomiebegriff in der Philosophie entwickelt hat. Das Hauptaugenmerk liegt auf der aristotelischen Tugendethik⁵ sowie dem moralphilosophischen Autonomiebegriff Immanuel Kants (1724-1804). In Kapitel 3 erläutere ich den Gebrauch des Autonomiebegriffes als Attribut bestimmter neuer Technologien, wie etwa autonome Systeme oder autonome Roboter. Anschließend wird in Kapitel 4 aufgezeigt, dass Autonomie nur dem Menschen zukommt: Aus moralphilosophischer Sicht ist dem bestehenden Autonomiebegriff durch autonome Maschinen nichts hinzuzufügen. Gleichzeitig ließe eine bestimmte Blickrichtung der Philosophie des Geistes eine Negation der Autonomie der Maschine zu. In Kapitel 5 soll der Versuch unternommen werden, die Autonomie der Maschine als Erweiterung der Autonomie des Menschen darzustellen.

Für eine grobe Unterscheidung benutze ich verallgemeinernd die Gegensatzpaare Mensch und Maschine, beziehungsweise Mensch und Technik. Spezifische Unterschiede von Technologien werden in Kapitel 3 herausgestellt.

An dieser Stelle weise ich darauf hin, dass sämtliche Fragen lediglich im Zusammenhang von Philosophie und Künstlicher Intelligenz erörtert werden, wobei auch die historische Entwicklung eine Rolle spielt. In der Forschung und Entwicklung ist der Gebrauch des Terminus „Künstliche Intelligenz“ (KI) oder „artificial intelligence“ (AI) mittlerweile zwar veraltet, da KI neben Psychologie, Philosophie, Linguistik, Neurowissenschaft und Anthropologie als ein Teilbereich der Kognitionswissenschaften angesehen wird.⁶ In dieser neuen Disziplin geht die KI vollständig auf, während die anderen Wissenschaften mit ihr lediglich Berührungspunkte haben. Diese Tatsache ist von Bedeutung für den Verlauf dieser Arbeit, da philosophische Fragen um den Autonomiebegriff über die Kognitionswissenschaften hinausgehen, innerhalb ihres Teil-

⁵ Aristoteles, 384-322 v. Chr., gilt als Begründer der Ethik als philosophische Disziplin.

⁶ K.wissenschaften bilden die Grundlage für die Entwicklung informationsverarbeitender Systeme (wobei die einzelnen Teildisziplinen unterschiedlich stark hervortreten), doch sind hierbei zum Teil auch andere Disziplinen wie Elektrotechnik oder Maschinenbau involviert. Da die entstehenden Technologien sehr unterschiedlich sind, werden sie daher nach Zweck und Beschaffenheit benannt: etwa als autonome, kognitive oder ubiquitäre Systeme, aber auch (humanoide) Robotik. Siehe Kap. 3.1.2.

bereichs KI der Autonomiebegriff jedoch einer eigenen Definition unterliegt. Jene Fragen, die bezüglich neuer Technologien etwa an der Schnittstelle zur Psychologie, Linguistik oder Sprachphilosophie entstehen, werden im Rahmen dieser Arbeit nicht behandelt.

Des Weiteren verzichte ich auf Bezüge zu Science Fiction in Literatur und Film, in denen Utopien und Dystopien autonomer Maschinen und Menschmaschinen gezeichnet werden.

*Nicht zehrender Krankheit erlagst du
empfindest nicht des Schwertes blutigen Lohn,
du lebst nach eigenem Gesetz, drum allein
zum Lande der Toten gehst du*

(Sophokles: Antigone, 819-822)

2. Der Autonomiebegriff in der Philosophie

Der Autonomiebegriff hat sich in der Philosophie in verschiedene Richtungen ausgeprägt. Die folgenden Unterkapitel sind historisch-chronologisch geordnet und orientieren sich vorwiegend an den Autonomiedefinitionen des sich autonom denkenden sowie dem autonom handelnden Menschen, die für den Fortgang der Arbeit relevant sind. Der bedeutende Einfluss des Autonomiegedankens in der politischen Philosophie kann hier nur angedeutet werden. Auch die Autonomiebegriffe anderer Disziplinen werden lediglich angesprochen, sofern sie sich auf das Mensch-Technik-Verhältnis beziehen.

2.1. Der Autonomiebegriff

Der aus dem Altgriechischen stammende Begriff „Autonomie“ (griech. *autonomia*, lat. *autonomia*) setzt sich aus den Begriffsteilen *autós* (selbst) und *nomós* (Gesetz) zusammen. Ins Deutsche ist der Begriff mit Selbstbestimmung, Selbstgesetzgebung sowie Eigengesetzlichkeit übersetzbar.⁷

2.2. Der Autonomiebegriff in der Antike

Im antiken Griechenland ist der Begriff zunächst mit einem politischen Sinn behaftet. Etwa Mitte des 5. Jahrhunderts v. Chr. ist mit der Forderung griechischer Stadtstaaten nach Autonomie jene nach Selbstständigkeit gemeint, sowie das Recht auf Unabhängigkeit von anderen Mächten.⁸ So erläutert der Geschichtsschreiber Herodot (ca. 490-420 v. Chr.) den Begriff Autonomie etwas umfassender als „innere und äußere politische Freiheit“⁹, während der jüngere Historiker Thukydides (460-395 v. Chr.) die politische Selbstbestimmung als wesentliches Merkmal der Autonomie betont.¹⁰ Die Römer hingegen gebrauchen den Begriff *autonomia* kaum, römische Schriftsteller greifen eher auf poetische Umschreibungen zurück, wie etwa Livius¹¹, welcher den Sinngehalt von Autonomie mit „*potestas vivendi suis legibus*“¹² umschreibt.

Trotz der vorwiegend politischen Konnotation des Begriffs ist er vereinzelt bereits in der Antike in ethischen sowie ästhetischen Kontexten nachgewiesen. In dem Zitat, welches dieses Kapitel einleitet, klagt der Chor in Sophokles Tragödie *Antigone*, die Protagonistin habe *autonómos*, „nach eigenem Gesetz“, gelebt und gehandelt und so ihr Todesurteil selbst verschuldet.¹³ Es lohnt sich, einen näheren Blick auf den Vers zu

⁷ Rosemarie Pohlmann: Autonomie, in: HWPh Bd. 1 (A-C) 1971, 701

⁸ Pohlmann 1971, 701

⁹ Ebd.

¹⁰ Ebd.

¹¹ Titus Livius, ca. 59 v. Chr. – 19 n. chr., römischer Geschichtsschreiber.

¹² Pohlmann schreibt die Nutzung dieser Wendung sowohl Livius als auch Caesar zu; eine andere Umschreibung von Autonomie war bei jenen Schriftstellern auch „*potestas utendi suis legibus*“. Siehe Pohlmann 1971, 702

¹³ Ebd.

werfen, denn neben dieser deutschen Übersetzung existieren auch andere, anhand derer sich die vielen Bedeutungsnuancen dessen ablesen lassen, was im Laufe der Zeit unter Autonomie verstanden und in der Literatur verarbeitet wurde:

Im Jahre 1804 übersetzt Friedrich Hölderlin den Vers mit: „Dein eigen Leben lebend, unter / Den Sterblichen einzig, / Gehst du hinab, in die Welt der Todten“¹⁴. In Claus Bremers Theaterstück *Antigone* sagt der Chor recht frei übersetzt: „du stirbst nicht durch das Schwert / du stirbst, weil du sterben willst“¹⁵. Bereits stark mit dem kantischen Verständnis von Autonomie konnotiert ist die auf Grundlage von Johann Donners¹⁶ Übersetzung entstandene Bearbeitung von Elke und Uwe Lehmann für das Hamburger Leseheft: „[...] Noch traf dich ein Schwert, das Rache gezückt; / Du schreitest freiwillig zum Hades.“¹⁷ Die Übersetzungen variieren somit vom Aufgreifen des antiken Konflikts zwischen staatlichem (d. h. menschengemachtem) und göttlichem Gesetz bis hin zu einer modernen Interpretation von Autonomie als ethischem Prinzip menschlichen Handelns. Inwieweit jedoch Sophokles selbst beabsichtigt hat, der Antigone mit dem Begriff *autonómos* mehr als eine politische Handlungsmotivation zuzuschreiben, sei dahingestellt. Auffällig sind dennoch die mannigfaltigen Übersetzungsmöglichkeiten, die Einflüsse aufweisen von der aristotelisch orientierten Handlungsethik hin zur kantischen Ethik der Privatautonomie.

2.3. Aristoteles‘ Ethik

Die Tugendethik des Aristoteles gilt als Gegenkonzept kantischer Pflichtethik. In der heutigen Zeit, da durch die Forschungsliteratur auch in Kants Ethik Lücken und Widersprüche bestätigt werden, erfährt die aristotelische Ethik, einst durch Kants Theorie der Autonomie als nicht mehr zeitgemäß diskreditiert, ein Wiederaufleben.¹⁸ Ich werde nun auf jene wesentlichen Passagen aus der *Nikomachischen Ethik*¹⁹ (NE) näher eingehen, welche für die Theoriebildung zur Plausibilisierung von autonomen Handlungen bis heute als Bezugspunkte herangezogen werden.

Zunächst nehme ich einen wesentlichen Punkt vorweg: Den Begriff „Autonomie“ hat Aristoteles selbst nie benutzt, die zentralen Begriffe seiner Ethik sind *eudaemonia*, das Glück, und die *areté*, Tugend. Aristoteles definiert Ethik als praktische Wissenschaft, die sich, im Gegensatz zu den sogenannten theoretischen und poetischen Wissen-

¹⁴ Sophokles / Friedrich Hölderlin / Claus Bremer: *Antigonae/Antigone*. 2 Texte 2 Inszenierungen 1 Theaterabend, Steinbach / Gießen 1969, 33. Hölderlins (1770-1843) Übersetzung des Dramas stammt aus dem Jahre 1804, während die Bearbeitung Bremers in den sechziger Jahren des 20. Jahrhunderts entstand.

¹⁵ Ebd.77

¹⁶ Johann Jakob Christian Donners (1799-1875) Übersetzung des Dramas erschien 1839.

¹⁷ Sophokles / Elke Lehmann: *Antigone*, ein Trauerspiel. Ungekürzter Text, auf Grundlage der Übersetzung von Donner, Husum 1995, 811f

¹⁸ Otfried Höffe: *Ausblick: Aristoteles oder Kant – wider eine plane Alternative*, in: ders. (Hg.): *Aristoteles: Die Nikomachische Ethik*, Berlin 1955, 277

¹⁹ Aristoteles: *Nikomachische Ethik*, auf der Grundlage der Übersetzung von Eugen Rolfes, hrsg. von Günther Bien, 4. durchges. Aufl., Hamburg 1985

schaften, mit dem Handeln des Menschen beschäftigt, welches genau dann für richtig gilt, wenn die Art der Handlung durch eine richtige Entscheidung hervorgerufen wurde.²⁰ Es geht also nicht um die Erkenntnis der Richtigkeit einer Handlung an sich, etwa einer tugendhaften, sondern um ihre Ausübung um eines guten Lebens Willen. Wie jedoch bestimmt wird, was beispielsweise als tugendhaft gilt oder was unter einem guten Leben zu verstehen ist, das sind Fragen, welche die Ethik als praktische Wissenschaft behandelt.²¹ Aristoteles definiert das gute Leben über die Ziele, die der Mensch verfolgt:

Jede Kunst und jede Lehre, desgleichen jede Handlung und jeder Entschluß, scheint ein Gut zu erstreben, weshalb man das Gute treffend als dasjenige bezeichnet hat, wonach alles strebt. [...] Wenn es nun ein Ziel des Handelns gibt, das wir seiner selbst wegen wollen, und das andere nur um seineswillen, und wenn wir nicht alles wegen eines anderen uns zum Zwecke setzen [...] so muß ein solches Ziel offenbar das Gute und das Beste sein.²²

Ein solches Ziel sieht Aristoteles etwa in der „Staatskunst“.²³ Somit lassen sich ihm zufolge die Ziele guten Handelns ausschließlich in einer Polis verwirklichen, da ihr Funktionieren gutes Handeln voraussetzt,²⁴

[...] da alles Wissen und Wollen nach einem Gute zielt, [...] welches als das Zielgut der Staatskunst bezeichnet werden muß, und welches im Gebiete des Handelns das höchste Gut ist.²⁵

Das teleologische Ziel, das die Menschen mit ihren Handlungen verfolgen, ist die Glückseligkeit (griech. *eudaemonia*): „Gut-Leben und Sich-gut-Gehaben [gilt] mit Glückselig-Sein als eins.“²⁶ Weiter heißt es:

Also: die Glückseligkeit stellt sich dar als ein Vollendetes und sich selbst Genügendes, da sie das Endziel alles Handelns ist.²⁷

Eudämonie soll als strukturierendes Ganzes verstanden werden, das die gesamte Lebensführung prägt. Doch Aristoteles verweist auf die voneinander abweichenden Vorstellungen von Glück unter verschiedenen Personengruppen. Zu diesen gehören (a) jene, die dem Genussleben (*bios apolaustikos*) frönen und die Lust zum Ziel haben, (b) jene,

²⁰ Vgl Jan Rohls: Geschichte der Ethik, Tübingen 1991, 64

²¹ Vgl. Rohls 1991, 64f

²² NE I 1, 1094a 1-4, 18-23

²³ Der Staatskunst als Wissenschaft unterstehen die anderen praktischen Wissenschaften Strategik, Ökonomik und Rhetorik. Sie alle sind zielgerichtet auf ein höheres Gut, nämlich dem der Staatsführung, welche ihnen wiederum „autoritativ vorschreibt, was man zu tun und was man zu lassen hat.“ (NE I 1, 1094b 6f)

²⁴ Siehe NE I 1, 1094b sowie NE I 2, 1095a, vgl. hierzu auch Rohls 1991, 64

²⁵ NE I 2, 1095a 15f

²⁶ NE I 2, 1095a 18ff

²⁷ NE I 5, 1097b 21ff

die einen praktisch-politischen Lebensstil (*bios politikos / praktikos*) verfolgen und auf Tugend beziehungsweise Ehre abzielen, sowie (c) jene, die mit ihrer theoretischen Lebensart zur Weisheit (*bios theôrêtikos*) gelangen wollen, wie etwa Philosophen.²⁸ Die Lebensform, die der Mensch für sich auswählt, kann jedoch auch falsch sein, sowohl in ihrer Ausübung als auch in ihrem Ziel. So ließe sich das subjektiv als geglückt oder gelungen empfundene Leben aus objektiver, das Ganze betrachtender Perspektive durchaus als missglückt einstufen.²⁹ Daher präsentiert Aristoteles in der *Nikomachischen Ethik* Kriterien, nach welchen das Glück, in Form eines gelungenen Lebens, sowohl subjektiv als auch objektiv bewertet werden kann.³⁰ Um diese näher zu erläutern, werde ich, dem Beispiel Jörn Müllers folgend, zunächst auf (1) die „formalen Kriterien“ und anschließend (2) auf die „materiale bzw. inhaltliche Bestimmung“ des aristotelischen Glücksbegriffs eingehen.³¹

(1) Aristoteles kennzeichnet die Eudämonie als *telos teleiôtaton*, das Ziel aller Ziele. Es ist das Gut, „das stets um seiner selbst und nie um eines anderen willen erstrebt wird“³², das für „sich selbst genügend“ (*autarkes*) und das „Allerbegehrteste“ (*hairêtôtaton*) ist.³³ Daher können aus diesem formalen Kriterium heraus die unterschiedlichen Lebensformen schon rein logisch danach eingeteilt werden, ob durch sie die Eudämonie erreicht wird oder nicht. Beispielsweise kann mit einem Leben, das lediglich auf den Gelderwerb abzielt, nicht das höchste Glück erreicht werden, wenn Geld bloß als Mittel zum Zweck, zum Erwerb anderer Zwecke, d.h. bestimmter Güter, dient.³⁴ Materielle, soziale und körperliche Faktoren wirken sich zwar begünstigend für die Lebensführung aus, sind jedoch nicht hinreichend notwendig zum Erreichen der Eudämonie.³⁵ Vielmehr kommt es darauf an, dass die genannten Faktoren in einem angemessenen Verhältnis zur Verfügung stehen, um sich positiv auf die Glückseligkeit eines Menschen auswirken zu können. Glück ist durch äußere Faktoren weder erweiterbar oder zu minimieren: Zwar ist Glück an sich schon hinreichend für ein gelungenes Leben, doch kann es weder durch Güter vergrößert werden, noch wird es durch sie „wählenswerter“. Glück ist das größte wählbare Gut, dem nichts hinzugefügt

²⁸ NE I 3, 1095b-1096a

²⁹ Vgl. NE I 3, 1095b

³⁰ Vgl. hierzu Jörn Müller: Glück als Vollendung menschlicher Natur. Die eudaimonistische Tugendethik des Aristoteles, in: Hanns-Gregor Nissing / Jörn Müller (Hgg.): Grundpositionen philosophischer Ethik. Von Aristoteles bis Jürgen Habermas, Darmstadt 2009, 26f

³¹ Vgl. ebd.

³² Ebd., 27

³³ NE I 5, 1097b 14, 16-23. Vgl. ebenso Christoph Horn: Antike Lebenskunst. Glück und Moral von Sokrates bis zu den Neuplatonikern, München 1998, 80

³⁴ Vgl. NE I 3, 1096a

³⁵ Horn 1998, 79

werden kann und das nicht mit anderen Gütern verrechenbar ist, da es „bereits alles Wertvolle einschließt und nicht verbesserungsfähig ist.“³⁶

(2) Generell gilt für Güter, dass sie entweder instrumentell oder um ihrer selbst willen gewählt werden, jedoch führen nur letztere zum Glück und sind somit wählenswert. Glück, so Aristoteles, ist eine dem Menschen spezifische Tätigkeit, da er sich hierdurch von der Tier- und Pflanzenwelt abhebt. Er begründet dies mit dem *ergon*-Argument: Jede Tätigkeit bringt ein eigentümliches Werk hervor, welches das Ziel der Tätigkeit ist.³⁷ Aristoteles verdeutlicht das *ergon*-Argument am Beispiel des Flötisten, der um des Flötenspiels willen spielt. Hieraus ergibt sich die Frage, worin denn das „eigentümliche Werk“ des Menschen als Mensch besteht. Glückseligkeit etwa zeichnet sich weder durch das Haben noch durch das Bekommen eines Werkes durch eine Tätigkeit aus, woraus Aristoteles ableitet, dass in einer Tätigkeit selbst die Glückseligkeit des Menschen liegen muss. Da das bloße Lebendigsein an sich auch den Pflanzen zukommt und selbst Tiere Sinnenwesen sind, „bleibt also nur ein nach dem vernunft-begabten [sic!] Seelenteile tätiges Leben übrig“³⁸. Vernunft gilt demnach als allgemeines Unterscheidungsmerkmal zwischen dem Menschen und anderen Lebewesen. Daraus folgt für Aristoteles, dass der Mensch die Eudämonie nur durch Vernunft, bzw. vernünftige Handlungen anstreben kann. Gerade in diesem Streben unterscheidet sich theoretische, zur Glückseligkeit fähige, von praktischer Tätigkeit sowie dem Genussleben, welche keine Zwecke an sich selbst sind, sondern um eines Gutes willen ausgeübt werden. Aristoteles macht die inhaltliche Frage nach dem Glück also nicht nur davon abhängig, ob Güter und Ziele instrumenteller oder intrinsischer Art sind, sondern teilt letzteres auch nach „unterschiedliche[n] Dignitätsgrade[n]“³⁹ ein. Mit seinen Worten: „Tugenden der Seele [...] [sind] teils sittliche oder Charaktertugenden, teils Verstandestugenden.“⁴⁰

Zusammengefasst ergibt sich aus den Aspekten (1) und (2), dass für ein glückseliges Leben nicht nur Rahmenbedingungen wie Lebensdauer oder Gesundheit ausschlaggebend sind, sondern insbesondere „[...] der kontinuierliche Vollzug der höchsten Vernunfttätigkeit [...]“⁴¹ entscheidend ist. Dieser Punkt wiederum ist bis in die heutigen Mensch-Technik-Debatten hinein aktuell.

³⁶ Horn 1998, 81

³⁷ Vgl. ebd.

³⁸ NE I 6, 1098a 2f

³⁹ Horn 1998, 82

⁴⁰ NE VI 2, 1138b 35 – 1139a 1. Zu den Tugenden der theoretischen Vernunft zählen Wissen, Geist und Weisheit (*epistémé, nús, sophia*), zu denen der praktischen Vernunft Kunst und Einsicht (*techné* und *phronésis*). Die im Hinblick auf die Eudämonie wählenswertesten Verstandestugenden, d. h. jene mit dem höchsten „Dignitätsgrad“, sind die theoretische Tugend, *sophia*, sowie die Einsicht, *phronésis*, als praktische Tugend. Vgl. NE, I 6, VI 2, Rohls 1991, 72 sowie Horn 1998, 82

⁴¹ Rohls 1991, 67

2.4. Autonomie als Forderung der Philosophie der Aufklärung

Bis in die Neuzeit hinein findet der Begriff Autonomie selten Erwähnung. Ausnahmen bilden politische Traktate von Denkern der Renaissance, z. B. Niccolò Machiavelli (1469-1527), mit welchem „[d]er Beginn des autonomen politischen Denkens [...]“⁴² in dieser Zeit verbunden wird. Die Aufnahme des Autonomiebegriffs im juristischen Bereich folgt mit den konfessionellen Auseinandersetzungen um den Augsburger Religionsfrieden (1555), in Rahmen dessen mit Autonomie von protestantischer Seite nicht nur Freiheit im religiösen Sinn, sondern in Anlehnung auf die griechische Bedeutung des Autonomiebegriffs politische Selbstbestimmtheit beansprucht wird.⁴³ Seither wurde der rechtstheoretische Autonomiebegriff in der Philosophie kontrovers diskutiert; auch Christian Wolff⁴⁴ benutzt den Begriff Autonomie in seiner *Philosophia civilis sive politicae* (1756) noch im Sinne politischer Selbstbestimmung.⁴⁵

Der politische, beziehungsweise rechtswissenschaftliche Autonomiebegriff gewinnt in dieser Zeit an Relevanz. In der Philosophie der Aufklärung entwickelt sich schließlich die Autonomie des Subjekts zur Grundidee und Kernforderung. Die Entstehung dieser Forderung ist jedoch, wie Kreimendahl hervorhebt, nicht denkbar ohne die Herausbildung eines zentralen Konzepts der Philosophie des 17. Jahrhunderts: die Erhebung des Ich zur „[...] unhintergehbare[n] Letztinstanz aller Gewissheit“⁴⁶ und zum „Fundament der philosophischen Weltanschauung[...]“⁴⁷. Gemeint ist, dass sich in der Neuzeit ein ethischer Paradigmenwechsel vollzieht: Während beispielsweise in der aristotelischen Ethik die Polis Bedingung für das moralisch gute Handeln des Menschen ist, treten in der neuzeitlichen Philosophie als Prämissen der Moral Begriffe wie Freiheit und Determination hervor. Eine Erörterung des Freiheitsbegriffs beispielweise erübrigt sich für Aristoteles, da er die Polis als einen Personenverband freier Bürger definiert und somit die Ziele guten menschlichen Handelns gleich derer einer funktionierenden Polis sind – in der Neuzeit hingegen wird in der Philosophie kein äußerer Rahmen vorgegeben, an welchem sich die Ziele des Menschen zu orientieren haben, „[...] sondern die eigentliche moralische Leistung besteht nun in der Setzung eines Ziels aus

⁴² Michael Stolleis: Geschichte des öffentlichen Rechts in Deutschland, München 1988, 90

⁴³ Vgl. Pohlmann 1971, 702. Interessanterweise wird im Reallexikon für Antike und Christentum der Begriff Autonomie nicht aufgeführt. Siehe: Theodor Klauser (Hg.): Reallexikon für Antike und Christentum : Sachwörterbuch zur Auseinandersetzung des Christentums mit der antiken Welt, Bd I.: A und O – Bauen, Stuttgart 1950

⁴⁴ Christian Wolff, 1679-1754

⁴⁵ „Cavenda igitur est omnis invasio alterius civitatis, omnisque usurpatio iurium ipsius publicorum, uti autonomiae“ Wolff [1756] 1998, §485, zitiert nach Pohlmann 1971, 707

⁴⁶ Lothar Kreimendahl: Einleitung. Die Philosophie des 18. Jahrhunderts als Philosophie der Aufklärung, in: ders. (Hg.): Philosophen des 18. Jahrhunderts, Darmstadt 2000, 17

⁴⁷ Kreimendahl 2000, 17. Wegweisend für diese Entwicklung hin zur Konzentration der Philosophie auf den Menschen sind bspw. die Werke von Pierre Bayle, John Locke, Gottfried W. Leibniz und Spinoza. Vgl. insbes. Kreimendahl 2000, 3-16

Freiheit und um der Freiheit willen.“⁴⁸ So wird Autonomie, ein für eine angestrebte Staatsform konstitutiver Begriff, in der idealistischen Ethik zum Moralprinzip des einzelnen Menschen erhoben.⁴⁹

Pieper schreibt daher zu Recht: „Die Ethik der Neuzeit ist eine Ethik der Freiheit.“⁵⁰ Beruft man sich auf Kant, so darf es ebenfalls heißen, die Ethik der Neuzeit sei eine Ethik der Autonomie, denn Kant setzt Freiheit mit der Gesetzgebung des Willens und der Autonomie gleich. Im folgenden Kapitel erörtere ich nun gezielt das Kantische Konzept von Autonomie.

2.4.1. Autonomie bei Kant

Seit Immanuel Kants theoretischer Auseinandersetzung mit einer Ethik der Autonomie in der *Grundlegung zur Metaphysik der Sitten* (1785) (GMS) gewinnt der Autonomiebegriff an Relevanz für die Moralphilosophie. Auch in der *Kritik der reinen Vernunft* (1781, 1787) (KrV) sowie in der *Kritik der praktischen Vernunft* (1788) (KpV) setzt sich Kant mit der Bedeutung von Autonomie als Selbstgesetzgebung der Vernunft, beziehungsweise als Selbstbestimmung des einzelnen Menschen auseinander.

Generell kann Autonomie als „[...] Strukturprinzip der gesamten Kantischen Philosophie [...]“⁵¹ angesehen werden. Kant entwickelt seine Gedanken zur Autonomie bereits in der *Kritik der reinen Vernunft*, wobei in dieser frühen Schrift Autonomie als Attribut der theoretischen Vernunft⁵² zu verstehen ist und – obwohl Autonomie begrifflich in der KrV nicht auftaucht – insbesondere in der dritten Antinomie von tragender Bedeutung ist.⁵³ Mit der Thesis zur dritten Antinomie leitet Kant eine Diskussion zur Kausalität ein:

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 durch Freiheit zu Erklärung derselben anzunehmen notwendig.⁵⁴

Es kann an dieser Stelle keine ausführliche Erörterung der dritten Antinomie geleistet werden, vielmehr soll jene Thesis auf die folgenden Ausführungen zum Zusammenhang

⁴⁸ Annemarie Pieper: Einführung in die Ethik, Tübingen / Basel ⁴2000 165. Vor diesem Hintergrund entwickelten sich sowohl Kants pflichtethische Autonomie als auch das teleologisch ausgerichtete Autonomieprinzip des Utilitarismus.

⁴⁹ Vgl. Rohls 1991, 5

⁵⁰ Pieper ⁴2000, 165

⁵¹ Pohlmann 1971, 707

⁵² Pohlmann 1971, 707

⁵³ Folgt man Kants Ausführungen über die KrV in seinem *Opus Postumum*, so handelt es sich um Autonomie, wenn das transzendente Subjekt „autonomische Verbindung empirischer Vorstellungen der Erscheinung gemäß“ leisten kann; daneben existiert die „A. der Ideen, in so fern [sic!] sie ein selbstständiges Ganz in Gegensatz der Erfahrung ausmachen“. (*Opus postumum*, Akad.-A. [22], [445] bzw. [81], zitiert nach Pohlmann 1971, 708). Es ist also Autonomie im Sinne einer theoretischen Vernunft.

⁵⁴ KrV [B 472]

von Autonomie und Freiheit in der GMS vorbereiten, in der diese „andere Kausalität durch Freiheit“ mit dem Wirken von Freiheit und eigener Gesetzgebung des Willens und damit der Autonomie gleichzusetzen ist.⁵⁵ Von zentraler Bedeutung ist Autonomie für Kants praktische Philosophie. In der GMS erklärt Kant „[d]ie Autonomie des Willens als oberstes Prinzip der Sittlichkeit“⁵⁶, die sich in der Freiheit des Menschen als Vernunftwesen äußert und deren „Prinzip ein kategorischer Imperativ“⁵⁷ ist.

Autonomie des Willens ist die Beschaffenheit des Willens, dadurch derselbe ihm selbst (unabhängig von aller Beschaffenheit der Gegenstände des Willens) ein Gesetz ist. Das Prinzip der Autonomie ist also: nicht anders zu wählen, als so, daß die Maximen seiner Wahl in demselben Willen zugleich als allgemeines Gesetz mit begriffen sein.⁵⁸

Ein Wille, welcher sittlich gut ist, äußert sich also in Form eines kategorischen Imperatives, welchen er sich qua Autonomie selbst gibt; das Prinzip der Autonomie ist somit gleichzusetzen mit dem kategorischen Imperativ und damit dem sittlich guten Willen. Konkret fasst Kant den Inhalt des Willens als praktische Gesetzgebung wie folgt zusammen:

Der Wille wird also nicht lediglich dem Gesetze unterworfen, sondern so unterworfen, daß er auch als selbstgesetzgebend und eben um deswillen allererst dem Gesetze (davon er selbst sich als Urheber betrachten kann) unterworfen angesehen werden muss.⁵⁹

Aus diesem Zitat geht besonders deutlich die Bedeutung von Autonomie als Selbstgesetzgebung hervor, zu welcher der Mensch als moralisch-praktisches Vernunftwesen Kant zufolge fähig ist. Der autonome Mensch gibt sich die Gesetze seines Handelns kraft seines Willens selbst. Ferner weist Kant darauf hin, dass der Mensch „[...] die Kausalität seines eigenen Willens niemals anders als unter der Idee der Freiheit denken [...]“⁶⁰ kann. Die einzelnen Komponenten können nicht isoliert gedacht werden, es besteht eine „Art von Zirkel“⁶¹, denn „[...] Freiheit und eigene Gesetzgebung des Willens sind beides Autonomie, mithin Wechselbegriffe [...]“⁶². Das Problem, welches sich für Kant hinsichtlich der „Idee der Freiheit“ stellt, müsste aufgrund der Wechselbegrifflichkeit jedoch auch für die Autonomie gelten. Zwar kann „[...] der Mensch die Kausalität seines eigenen Willens niemals anders als unter der Idee der Freiheit denken [...]“⁶³. Dennoch würde Kant an die Grenzen der praktisch-

⁵⁵ Vgl. GMS [450]

⁵⁶ GMS [440]

⁵⁷ GMS [441]

⁵⁸ GMS [440]

⁵⁹ GMS [431]

⁶⁰ GMS [452]

⁶¹ GMS [450]

⁶² Ebd.

⁶³ GMS [452]

vernünftigen Erklärbarkeit der Idee der Freiheit gelangen,⁶⁴ stellte er sie nicht als das vernunftgemäß Denkbare der negativen Freiheit und Heteronomie gegenüber.⁶⁵ Im Gegensatz zu Autonomie bezeichnet Heteronomie den Zustand,

[w]enn der Wille irgend worin anders als in der Tauglichkeit seiner Maximen zu seiner eigenen allgemeinen Gesetzgebung, mithin, wenn er, indem er über sich selbst hinausgeht, in der Beschaffenheit irgendeines seiner Objekte das Gesetz sucht [...]⁶⁶

Heteronomie ist die „Naturnotwendigkeit [...] der wirkenden Ursachen“⁶⁷, wohingegen Autonomie „die Eigenschaft des Willens, sich selbst Gesetz zu sein“⁶⁸ ist. Dementsprechend handelt es sich um negative Freiheit, wenn – vergleichbar der „Naturnotwendigkeit [als] Eigenschaft der Kausalität aller vernunftlosen Wesen“⁶⁹ – der Wille als Kausalität vernunftfähiger Lebewesen gelten soll, wobei auch in diesem Fall Freiheit die „Eigenschaft dieser Kausalität“⁷⁰ sei. Positive Freiheit hingegen übersteigt den Kausalitätsbezug von Ursache und Wirkung der Sinnenwelt, indem sie auf die Selbstgesetzgebung des guten Willens rekurriert, wobei sie jedoch ein synthetischer Satz ist, der sich auf etwas Drittes bezieht, „[...] worauf uns die Freiheit weist und von dem wir *a priori* eine Idee haben [...]“⁷¹. Kant schließt daraus, dass „[...] jedem vernünftigen Wesen, das einen Willen hat, notwendig auch die Idee der Freiheit [...]“⁷² innewohnt und somit ist jeder Mensch zu positiver Freiheit fähig. Allein dadurch, dass der vernünftige Mensch denken kann, hält Kant es für abwegig, dass der Mensch seine Urteilskraft durch einen Antrieb anstatt durch seine Vernunft erklärt. Zwar unterliegt der Mensch sowohl der „Sinnenwelt“ als auch der „Vernunftwelt“, doch allein die „Unabhängigkeit von den bestimmenden Ursachen der Sinnenwelt [...] ist Freiheit.“ Diese Freiheit, die mit der Autonomie des freien Willens eine Einheit bildet, qualifiziert

⁶⁴ In der Fußnote zu GMS [448] weist er jedoch darauf hin, dass es hinreichend sei, von einer Idee der Freiheit zu sprechen, ohne diese theoretisch beweisen zu müssen, da für Kant der vernünftig handelnde Mensch bereits frei ist, wenn er sich in seinem praktischen Handeln an der Idee der Freiheit orientiert (also seine Autonomie als Selbstgesetzgebung seines Willens ausübt). Vgl. hierzu GMS [447f], insbes. das Kapitel „Freiheit muß als Eigenschaft des Willens aller vernünftiger Wesen vorausgesetzt werden“.

⁶⁵ Vgl. GMS [433] und insbes. [452f]

⁶⁶ GMS [444]. Siehe auch die Ausführungen zur Heteronomie in GMS [444]: „[...] der Wille gibt sich nicht selbst, sondern ein fremder Antrieb gibt ihm vermittelt einer auf die Empfänglichkeit desselben gestimmten Natur des Subjekts, das Gesetz.“

⁶⁷ GMS [446]

⁶⁸ GMS [447]

⁶⁹ GMS [446]

⁷⁰ Ebd.

⁷¹ GMS [447]. Was die Idee der Freiheit sein soll, lässt sich nur durch einen synthetischen Satz erklären, bzw. „[...]die Deduktion des Begriffs der Freiheit aus der reinen praktischen Vernunft, mit ihr auch die Möglichkeit eines kategorischen Imperativs, begreiflich [zu] machen, [...] bedarf noch einiger Vorbereitung.“ (GMS [447])

⁷² GMS [448]

den Menschen zum vernünftigen Handeln.⁷³ Wann das menschliche Handeln einen heteronomen, d.h. fremdbestimmten, beziehungsweise einen autonomen, d.h. selbstbestimmten Ursprung hat, hebt Kant in der KpV wie folgt hervor:

Die sinnliche Natur vernünftiger Wesen überhaupt ist die Existenz derselben unter empirisch bedingten Gesetzen, mithin für die Vernunft Heteronomie. Die übersinnliche Natur ebenderselben Wesen ist dagegen ihre Existenz nach Gesetzen, die von aller empirischen Bedingung unabhängig sind, mithin zur Autonomie der reinen Vernunft gehören.⁷⁴

Demnach ist Heteronomie der Handlungsraum des Menschen in der Sinnenwelt und Autonomie der Handlungsraum in der Vernunftwelt. Neben den bisher aufgeführten Inhalten von Autonomie des Verstandes, dann Autonomie der Vernunft, sei noch kurz auf die Heautonomie hingewiesen, welche Kant in der *Kritik der Urteilskraft* (KU) darlegt. Unter Heautonomie versteht Kant eine die Autonomie reflektierende Autonomie, die sich als Autonomie „der reflektierenden Urteilskraft für unseren subjektiven Vernunftgebrauch“⁷⁵ äußert. Ich fasse die Ausführungen zu Kants transzendental- und insbesondere moralphilosophischem Autonomiebegriff an dieser Stelle mit einem Zitat Kants aus der ersten Einleitung in die KU zusammen, welches die verschiedenen Prägungen von Autonomie bündelt:

[...] diese Autonomie [(gemeint ist Heautonomie) (Anm. d. Verf.)] aber ist nicht (so wie des Verstandes, in Ansehung der theoretischen Gesetze der Natur, oder der Vernunft, in practischen Gesetzen der Freiheit) objectiv, d. i. durch Begriffe von Dingen oder möglichen Handlungen, sondern bloß subjectiv, für das Urtheil aus Gefühl gültig, welches, wenn es auf Allgemeingültigkeit Anspruch machen kann, seinen auf Principien a priori gegründeten Ursprung beweiset.⁷⁶

Nachdem Kant in seiner Transzendental- und Moralphilosophie die Funktion von Autonomie bestimmt hat, erörtert er in seinen späteren Monographien *Metaphysik der Sitten* (MdS) (1797) und *Streit der Fakultäten* (1798) den historischen Autonomiebegriff, welcher die politische sowie institutionelle Selbstbestimmung zum Inhalt hat. So erläutert Kant in der MdS das Wesen eines freien Staates wie folgt:

Also sind es drei verschiedene Gewalten (potestas legislativa, executorial, iudiciaria), wodurch der Staat (civitas) seine Autonomie hat, d.i. sich selbst nach Freiheitsgesetzen bildet und erhält.⁷⁷

⁷³ Siehe GMS [452f]. Kant führt an jener Stelle zum Vergleich an, dass dem entgegengesetzt auf Ebene der Sinnenwelt das Naturgesetz den Erscheinungen zugrunde liegt.

⁷⁴ KpV [74]

⁷⁵ Pohlmann 1971, 708. Vgl. hierzu auch Maximilian Forschner: *Gesetz und Freiheit. Zum Problem der Autonomie bei I. Kant*, München / Salzburg 1974, 134 u. 270ff

⁷⁶ KU, Erste Einleitung in die KU, VIII [225]

⁷⁷ MdS [318]

Dem institutionellen Autonomiebegriff liegt dennoch das Verständnis von Autonomie durch Vernunft zugrunde, schließlich sind laut Kant die Vertreter autonomer Staaten und Institutionen stets vernünftige und somit autonome Personen, die kraft ihres gemeinsamen vernünftigen Willens autonome Staatsbürger sind und somit die Basis eines autonomen Staates bilden.⁷⁸

Basierend auf Kants umfassendem Autonomiebegriff entstanden im Laufe der Zeit differenzierte Konzepte der Autonomie. Wie noch gezeigt wird, sind diese jedoch zunehmend in einen teleologischen Rahmen gefasst – sofern sie nicht die kantische Autonomiedefinition negieren.

2.3.2. Reaktionen auf Kants Autonomieverständnis

Die oben vorgestellte kantische Auffassung von Autonomie löste in der Philosophie des 19. Jahrhunderts zunächst eine kontroverse Diskussion des Begriffes aus, fand im Laufe der Zeit jedoch eine wachsende Zustimmung und wurde teilweise erweitert. Somit erfolgte die Auseinandersetzung mit dem Autonomiebegriff Kants auf drei Ebenen, die sich teilweise zeitlich und inhaltlich überlappen.

(1) Als erstes sei die gegnerische, theologisch geprägte Position erwähnt: Jacobi⁷⁹, Schlegel⁸⁰ und Reinhold⁸¹ kritisierten Kants auf Selbstbestimmung durch Vernunft basierenden Autonomiebegriff, da dieser keinen Raum für die Erklärung des göttlichen Wirkens lasse.⁸²

(2) Dementgegen wurde in der kritischen und spekulativen Philosophie das Prinzip der Autonomie der Vernunft aufgegriffen, um „[...] es in seine Konsequenzen zu entwickeln [...]“.⁸³ Das heißt: Zwar bleibt der Grundgedanke moralischer Autonomie weiterhin präsent, jedoch wird er aus der kantisch-deontologischen Theorie herausgegriffen und in den aufkommenden teleologisch-normativen Theorien ausgearbeitet.⁸⁴ Johann Gottlieb Fichte⁸⁵ (1762-1814) führt beispielsweise den Autonomiegedanken

⁷⁸ Vgl. Pohlmann 1971, 709. Dementsprechend fußt bspw. auch die „A[utonomie] der Universität auf der wissenschaftlichen Denkungsart ihrer Gelehrten“. (Vgl. ebd.)

⁷⁹ Friedrich Heinrich Jacobi, Philosoph und Jurist, 1743-1819.

⁸⁰ Karl Wilhelm Friedrich von Schlegel, Philosoph und Schriftsteller, 1772-1829.

⁸¹ Karl Leonhard Reinhold, österreichischer Philosoph, 1757–1823.

⁸² Vgl. Dieter Henrich: Selbstverhältnisse, Stuttgart 1993, 42f sowie Pohlmann 1971, 709f

⁸³ Henrich 1993, 43

⁸⁴ Vgl. Annemarie Pieper: Autonomie, in Wilhelm Korff / Lutwin Beck / Paul Mikat (Hgg.): Lexikon der Bioethik Bd. 1, Gütersloh 1998, 290

⁸⁵ Von Kants Philosophie beeinflusst ist insbesondere Fichtes Vorstellung von der „absolute[n] Existenz und Autonomie des Ich“¹ als Grundlage jenes Handelns, das dem Bewusstsein zugrunde liegt. Dem kantischen Vorbild entsprechend definiert Fichte folgende drei Bedeutungen von Autonomie: (1) „eine ununterbrochene Gesetzgebung an sich selbst“; (2) „absolute Unbestimmbarkeit durch etwas ausser dem Ich“ und (3) „absolut freie Reflexion des Ich auf sich selbst“². (¹ zit. nach Pohlmann 1971, 709; ² Johann Gottlieb Fichte: System der Sittenlehre nach den Principien der Wissenschaftslehre (1798), in: I. H. Fichte (Hg.): Fichtes Werke, Bd. IV: Zur Rechts- und Sittenlehre, Nachdruck der Ausgabe Berlin 1845/46, Berlin 1971, 56f. Vgl. auch Pohlmann 1971, 709)

Kants weiter, indem er die absolute Autonomie der Vernunft als Ausgangspunkt jeglicher Philosophie definiert:

Entweder, alle Philosophie muss aufgegeben werden, oder die absolute Autonomie der Vernunft muss zugestanden werden.⁸⁶

Auch Hegel (1771-1831) und Schelling (1775-1814) sind dieser Tradition der „absoluten Autonomie der Vernunft“⁸⁷ verpflichtet. Jedoch ist in Hegels frühen Schriften Autonomie ohne die christliche Tradition nicht denkbar; in seiner Religionsphilosophie entwickelt er schließlich den Gedanken der absoluten Autonomie dahingehend weiter, dass „das Christentum die einzig wahre und absolute Religion sei.“⁸⁸ Schelling betrachtet Autonomie hingegen aus dem Gesichtspunkt der Ästhetik heraus und argumentiert, „absolute Autonomie“ käme nur dem ästhetischen Genie zu.⁸⁹ Genauso setzt sich bereits Friedrich Schiller (1775-1805) mit der Autonomie des Ästhetischen auseinander, wobei er diese nicht wie Schelling aus der theoretischen Vernunft herleitet, sondern sich mehr für das „ästhetische Urteilsvermögen“ interessiert.⁹⁰

(3) Jedoch gab es auch stets Tendenzen, sowohl in den Geistes- als auch in den Naturwissenschaften, jegliche Theorien von der Autonomie des Individuums zu unterminieren. Eine Welle konkreter Auseinandersetzungen mit vorherrschenden Autonomiedefinitionen fand an der Schwelle zum 20. Jahrhundert statt. Rohrmoser fasst die wesentlichen Positionen, die moralische Autonomie verneinen, wie folgt zusammen:

Autonomie kann nunmehr entlarvt werden als das Produkt einer nicht gelungenen Verdrängung faktischer Heteronomie, die das autonome Ich mit dem Verlust des Gottes, (Kierkegaard), mit dem Verlust des anderen als des rettenden Du (Feuerbach), dem Verlust der Geschichte als Tradition und dem Vergessen des Todes (Heidegger) oder gar mit Krankheit (Freud) oder mit dem Verlust seines wahren gesellschaftlichen Seins (Marx) bezahlen musste.⁹¹

Ebenfalls negiert wird moralische Autonomie in Theorien, in denen das menschliche Verhalten rein naturwissenschaftlich erklärt wird, z.B. indem man „moralisches Handeln ausschließlich mit den sprachlichen Mitteln der Evolutionstheorie rekonstruieren zu können“⁹² glaubt. Solche naturkausal hergeleiteten Erklärungen

⁸⁶ Fichte (1798) 1971, 59

⁸⁷ Ebd.

⁸⁸ Günter Rohrmoser: Autonomie, in: Herman Krings / Hans Michael Baumgartner / Christoph Wild (Hgg.): Handbuch philosophischer Grundbegriffe, Bd. 1 Das Absolute – Gesellschaft, München 1973, 158

⁸⁹ Ebd., 158

⁹⁰ Vgl. hierzu Pohlmann 1971, 710

⁹¹ Rohrmoser 1973, 158

⁹² Pieper 1998, 290

menschlichen Handelns lassen keinen Raum für ethisch-moralische Begriffe wie Autonomie, Freiheit oder Menschenwürde.

2.5. Ausdifferenzierung des kantischen Autonomiebegriffs bis in die heutige Zeit

Vor dem Hintergrund dieser bis ins 20. Jahrhundert hineinreichenden Diskurse wurde der Begriff Autonomie zu einem Synonym für Willensfreiheit oder allgemein zu einem Freiheitsbegriff im ethischen Sinne.⁹³ Ab dem Neukantianismus erfolgte die erneute Rezeption des kantischen Verständnisses von Privatautonomie als ethischem Begriff, ebnete jedoch gleichzeitig den Weg zur Kontextualisierung und Neudefinition des Autonomiebegriffes innerhalb neuer Wissenschaftsdisziplinen. Dass Autonomie, wie sie in der Philosophie bestimmt wurde, sei es als politische oder moralische, für unser heutiges Verständnis vom Menschen weiterhin von zentraler Bedeutung ist, zeigt sich an seiner Relevanz in geisteswissenschaftlichen Teildisziplinen. So haben sich seit dem Neukantianismus fünf wesentliche Geltungsbereiche der Autonomie herausgebildet: Die Autonomie der Wissenschaft, des Sozialen, der Religion, des Ästhetischen sowie die sittliche Autonomie.⁹⁴ Nachfolgend werde ich auf einige neukantianische Positionen eingehen, welche sich für das Autonomieverständnis des 20. Jahrhundert als konstitutiv hervorgetan haben.

Stellvertretend für eine ganze Reihe von Neukantianern wie Natorp, Görland und Liebert sei hier auf die Arbeit Hermann Cohens verwiesen, welcher in seiner *Ethik des reinen Willens* (1904, 1907) mit der historischen Darstellung der Autonomie des Selbstbewusstseins befasst, die sich schließlich in den vier Formen Selbstgesetzgebung, Selbstbestimmung, Selbstverantwortung und Selbsterhaltung niederschlägt.⁹⁵ Somit wird die moralische Autonomie ebenso zur Grundlage anderer Disziplinen und als „allgemeines philosophisches Prinzip“⁹⁶ anerkannt. Ebenso bezeichnet Johannes Volkelt in seinem *System der Ästhetik*⁹⁷ die „vier Betätigungsweisen“⁹⁸, das „wissenschaftliche Erkennen“⁹⁹, das „sittliche Wollen“¹⁰⁰, das „religiöse und

⁹³ Vgl. Pohlmann 1971, 711

⁹⁴ Vgl. Ebd.

⁹⁵ Vgl. Hermann Cohen: *Ethik des reinen Willens*, Nachdr. der 2. revidierten Aufl. Berlin 1907, Hildesheim 1981, 324-388

⁹⁶ H. Rickert: *Allgemeine Grundlegung der Philosophie* (1921), 310, zit. nach Pohlmann 1971, 712

⁹⁷ Johannes Volkelt: *Kunstphilosophie und Metaphysik der Ästhetik*, München²1924

⁹⁸ Volkelt 1924, 545

⁹⁹ Ebd. „Das Ausgezeichnete des wissenschaftlichen Erkennens liegt darin, daß der wissenschaftliche Erkennende seinen Gedanken mit Autonomie gegenübersteht. [...] Indem ich meinem Denken autonom gegenüberstehe, bin ich auch des Gegenstandes, auf den sich mein Denken bezieht, theoretisch Herr.“

¹⁰⁰ Ebd., 456: „Auch das Auszeichnende des sittlichen Wertes liegt in der Autonomie des Individuums. Nur ist es hier nicht die Autonomie hinsichtlich der logischen Vorstellungsverknüpfungen, sondern hinsichtlich des Sollens und Wollens.“

künstlerische Verhalten“¹⁰¹ als autonom, da diese allesamt der „Autonomie des Individuums“¹⁰² entspringen. Diese Definition von Autonomie beschränkt sich auf das Autonom-Sein des Einzelnen, sowie auf dessen Handlungen.

In den jüngeren wissenschaftlichen Disziplinen wie Soziologie, Pädagogik und Psychologie wurden bisweilen eigene Definitionen von Autonomie etabliert, die das autonome Verhalten des Menschen in einen gesellschaftlichen Kontext setzen. In der Soziologie hat sich etwa die Unterscheidung zwischen funktionaler und personaler Autonomie etabliert.¹⁰³ Letztere äußert sich in „Individualität, Spontaneität, persönlicher Selbstbestimmung usw. [und] wird vorwiegend in dem Kontext ihrer Bedrohung durch gesellschaftlichen Zwang diskutiert“¹⁰⁴. In der Soziologie findet der Autonomiebegriff als erstes bei Max Weber Erwähnung, welcher unter Autonomie ein funktionales Prinzip begreift, das der Heteronomie, d.h. Ordnungsinstanzen Außenstehender, gegenübersteht:

Denn der Begriff der Autonomie ist [...] an das Bestehen eines nach Merkmalen, sei es auch wechselnden, jeweilig irgendwie abgrenzbaren Personenkreis geknüpft, welcher kraft Einverständnis oder Satzung einem von ihm prinzipiell selbstständig abänderbaren Sonderrecht untersteht.¹⁰⁵

Autonomie als funktionales Prinzip äußert sich demnach in der Selbstgesetzgebung eines Personenkreises, etwa eines Vereins, einer Aktiengesellschaft oder Gemeinde. Niklas Luhmann definiert in seinem Buch *Zweckbegriff und Systemrationalität* (1968) Autonomie als systemstrukturelles Prinzip:

Autonomie ist nicht in Kausalkategorien als ursachlose Spontaneität zu begreifen, sondern nur systemstrukturell als Selbstprogrammierung. Sie besteht darin, daß das System sich durch eigene Programme in die Lage versetzt, an beiden Zeitgrenzen, sowohl im Hinblick auf die Ursachen als auch im Hinblick auf die Wirkungen seines Handelns, Informationen der Umwelt aufzunehmen, selektiv zu verarbeiten.¹⁰⁶

Als Gesellschaftstheoretiker bezieht sich Luhmann zwar auf ein System, in das Menschen eingebunden sind, wie etwa ein politisches System. Er überträgt die Definition personaler Autonomie auf die funktionale, die in Gesellschaften zum Tragen kommt. Auffällig ist dennoch die techniklastige Sprache: Selbstprogrammierung, Informationen verarbeiten – allein aus der Definition ist der Bezug zur autonomen

¹⁰¹ Volkelt 1924, 547: „Autonom sind auch das religiöse und das künstlerische Verhalten [...]. Der völlig mündig gewordene Mensch zollt auch auf diesen beiden Gebieten nur dem aus selbstgesetzgeberischer Betätigung Gewonnenen Anerkennung.“

¹⁰² Ebd., 546

¹⁰³ Vgl. Pohlmann 1971, 718

¹⁰⁴ Ebd. Pohlmann bezieht sich hierzu auf Herbert Marcuses Schrift „Der eindimensionale Mensch“.

¹⁰⁵ Max Weber: *Wirtschaft und Gesellschaft*, hrsg. von Werner Gephart und Siegfried Hermes, Tübingen 2010, 368

¹⁰⁶ Niklas Luhmann: *Zweckbegriff und Systemrationalität*, Tübingen 1977, 104f

Maschine ebenfalls möglich, denn dass in Luhmanns System Menschen agieren, wird erst aus dem weiteren Kontext ersichtlich. Luhmanns Definition zeigt im Übrigen deutliche Parallelen zur aristotelischen Handlungsmotivation, nach welcher die Polis Rahmenbedingung der Eudämonie ist.

Jürgen Habermas, welcher Autonomie eher diskursgebunden als aprioristisch denkt und diese sowohl als politischen als auch moralischen Begriff erörtert, führt personale und funktionale Autonomie im Diskurs zusammen. Er distanziert sich von dem kantischen Autonomieverständnis, welches das Demokratieprinzip aus der moralischen Selbstgesetzgebung ableitet:

Die Idee der Selbstgesetzgebung von Bürgern darf also nicht auf die moralische Selbstgesetzgebung einzelner Personen zurückgeführt werden. Autonomie muss allgemeiner und neutraler begriffen werden. Deshalb habe ich ein Diskursprinzip eingeführt, das gegenüber Moral und Recht zunächst indifferent ist.¹⁰⁷

Habermas führt das Diskursprinzip ein, um über dieses eine plausible, angewandte Herleitung der Entstehung des Demokratieprinzips leisten zu können.¹⁰⁸ Autonomie ist zwar ein grundlegendes Merkmal menschlichen Handelns und Verhaltens, muss jedoch nicht zwangsläufig mit Moralität erklärt werden, sondern lässt sich daran ablesen, dass der Mensch aufgrund seiner Diskursfähigkeit selbstgesetzgebend handeln und etwa die Demokratie hervorbringen kann. Für das Autonomieverständnis an sich bedeutet dies, dass es erst in der diskursiven Reflexion innerhalb eines gesellschaftlichen Kontextes gefestigt wird und damit, je nach Umfeld, auch eine Neuausrichtung offenlässt. Ein weiterer Unterscheidungspunkt zu Kant ist die teleologische Begründung menschlichen Handelns. In seinem Hauptwerk *Theorie des kommunikativen Handelns* nennt Habermas „[...] die teleologische Struktur für *alle* Handlungsbegriffe fundamental.“¹⁰⁹

In der Ethik gilt Autonomie gegenwärtig als „[...] Kennzeichen des Humanen [und] ist ein unverlierbares Grundcharakteristikum, das jedem menschlichen Lebewesen zuerkannt wird [...]“.¹¹⁰ Dieses ethische Verständnis von Autonomie existiert unabhängig von der Möglichkeit und Fähigkeit des Einzelnen, seine Autonomie auszuleben oder sich ihrer überhaupt bewusst zu werden. Der Mensch besitzt personale Autonomie, weil er Mensch ist; die Bedingungen autonomen Handelns und Entscheidens an sich, wie Intentionalität und Zweckrationalität, sind nicht die der personalen Autonomie.¹¹¹ Dass jenen Menschen, die aufgrund von physischen und psychischen Handicaps „[...] keinen vernünftigen Gebrauch von ihrer Freiheit machen

¹⁰⁷ Jürgen Habermas: Faktizität und Geltung, Frankfurt a. M. 1992, 154

¹⁰⁸ Vgl. ebd. 154ff.

¹⁰⁹ Jürgen Habermas: Theorie des kommunikativen Handelns, Bd. 1 Handlungsrationalität und gesellschaftliche Rationalisierung, Frankfurt a. M. 1987, 151f

¹¹⁰ Pieper 1998, 291

¹¹¹ Ebd.

können [...]“¹¹², überhaupt personale Autonomie und Menschenwürde zukommt und gesichert wird, gelingt erst durch die Leistung und Zuschreibung derer, die auch autonom handeln, d.h. der „[...] Freiheit selbstverantwortlich Grenzen setzen [...]“¹¹³ können. Autonomie dient dabei als „moralische Kompetenz“¹¹⁴ und so entstehen – zumindest in demokratischen Rechtssystemen – Maßstäbe zur Beurteilung von Gut, Böse, Zurechnungsfähigkeit und Unzurechnungsfähigkeit. Jenes böse Handeln aber äußert sich im wissentlichen Missbrauch der eigenen Autonomie, um die Freiheit anderer zu beschneiden. In einem solchen Fall würde dem betreffenden Menschen zwar die individuelle Freiheit, nicht jedoch seine personale Autonomie entzogen werden können.¹¹⁵

2.6. Zusammenführung und Fortgang

Für den heutigen Begriff der Autonomie in der Philosophie bleibt festzuhalten, dass er keine allgemeinverbindliche Definition besitzt, sondern zunächst nach dem spezifischen Geltungsbereich entschieden werden muss, um den Bedeutungsinhalt näher bestimmen zu können. An dem aufgeführten historischen Abriss habe ich aufgezeigt, dass der Inhalt dessen, was unter Autonomie verstanden wird, nicht nur in der Vergangenheit stark an Kontexte geknüpft wurde, sei es Politik, Ethik, Ästhetik, Religion oder Psychologie.

Resümiert man die vorgestellten Definitionen, die sich in der Philosophie entwickelt haben, so fällt auf: Sie tragen allesamt die Grundzüge des kantischen Autonomieverständnisses vom selbstgesetzgebenden, sittlichen Individuum, der Rahmen seines autonomen Handelns ist jedoch stets von einem teleologischen Ziel umgeben. Die Autonomie eines Menschen kann jedoch im Konflikt zur Heteronomie stehen. Genauso muss der gesellschaftliche Kontext betrachtet werden, in dem die Autonomie des Menschen aber auch der Gesellschaft zum Tragen kommt.

Die zwei wesentlichen Positionen der Philosophie, autonomes Handeln – oder generell Handlungsmotivationen des Menschen – entweder teleologisch oder deontologisch zu motivieren, sind auch für den aktuellen Diskurs, der nun um den Bereich des Technischen erweitert wurde, weiterhin wegweisend. Welche der bereits existenten Autonomiedefinitionen noch taugen, um den sich immer mehr technifizierenden

¹¹² Pieper 1998, 291

¹¹³ Ebd.

¹¹⁴ Ebd.

¹¹⁵ Vgl. Pieper 1998, 291. Dennoch gibt es auch schwer entscheidbare Situationen, in denen die Autonomie des einen Menschen bereits die Autonomie des Anderen beschränkt. Ethische Fragen der Autonomie spielen daher eine wesentliche Rolle in der Medizin. In Kontroversen um die Reproduktionsmedizin und gentechnische Verbesserung des Menschen stehen sich etwa die Autonomie der sogenannten Designer und die des entworfenen, designten Menschen gegenüber. Vgl. hierzu auch Michael J. Sandel: Plädoyer gegen die Perfektion, aus dem amerikan. von Rudolf Teuwsen, Berlin 2008. Sandel warnt vor der Aushöhlung unseres Verständnisses vom Leben als „genetische Lotterie“ und den damit verbundenen Begriffen von Fairness, Ehrfurcht und sozialer Gerechtigkeit durch Gentechnik.

Menschen zu beschreiben, werde ich im weiteren Verlauf dieser Arbeit darlegen. Dies hängt wesentlich davon ab, welcher Status dem Menschen, und welcher der Maschine zugeschrieben wird. Im Hinblick auf autonome Maschinen stellt sich ebenfalls die Frage, inwiefern sie von sich aus bereits Bedingungen von Autonomie erfüllen oder auf Grundlage welcher Kriterien hier bereits von Autonomie ausgegangen wird.

Nevertheless I believe that at the end of the century the use of word and general educated opinion will have altered so much that one will be able so speak of machines thinking without expecting to be contradicted.

(Alan Turing: Computing Machinery and Intelligence, 442)

3. Autonomie der Technologien

In diesem Kapitel gehe ich zunächst der Frage nach, was eine autonome Maschine ist und wann sie als autonom gilt. Dabei zeigt sich, dass Autonomie in Bezug auf Maschinen als Attribut verstanden wird, das auf neue Möglichkeiten in ihrer Nutzung als Artefakt hinweist und sich damit von dem zuvor erläuterten philosophischen Autonomiebegriff unterscheidet. Die Autonomie der Maschine ist als ein Gegenstandsbereich der technologischen Entwicklung zu betrachten, welche einer eigenen Definition unterliegt. Hierbei gilt Autonomie nicht als Substantivierung des Attributs „autonom“, sondern ist als Autonomisierung, als eine Erweiterung der technischen Möglichkeiten der Mechanisierung zu verstehen. Einen besonderen Augenmerk lege ich auf jene Kriterien, die den fließenden Übergang von Automatisierung zu Autonomisierung forcieren: Kontrolle als unerlässliches Kriterium der Autonomisierung sowie Künstliche Intelligenz als Basis zur Entwicklung autonomer Technologien. Um zwischen Autonomie als philosophischen Begriff sowie Autonomie im technischen Bereich besser zu unterscheiden, wird im Folgenden von „menschlicher Autonomie“ oder „technischer Autonomie“ die Rede sein.

3.1. Begriff und Gegenstand

In diesem Kapitel untersuche ich zunächst, wie der Begriff „Autonomie“ im Kontext neuer Technologien definiert wird. Im zweiten Schritt wird der technologische Gegenstand selbst, der als autonom bezeichnet wird, näher betrachtet.

3.1.1. Das Attribut ‚autonom‘

Im Jahre 1986 fand in Amsterdam eine internationale Konferenz mit dem Titel „Intelligent Autonomous Systems“ statt, auf der theoretische sowie praktische Aspekte des Designs und Betriebs autonomer Systeme diskutiert wurden.¹¹⁶ Auf jener Konferenz wurde unter Autonomie folgender technologischer Fortschritt zusammengefasst:

[...] autonomy is obtained by partitioning of the task into three basic parts consisting of a sensor system, a control module and a world model.¹¹⁷

Die Definition des Autonomiebegriffs erfolgt über die Beschreibung jenes technologischen Levels, das erreicht werden muss, damit eine Maschine als autonom bezeichnet werden kann. Der Begriff Autonomie tritt im technischen Bereich vordergründig in Form des Adjektivs „autonom“ auf, etwa als Attribut von sogenannten Agenten oder Systemen. George Bekey erklärt den Begriff Autonomie in Bezug auf Systeme folgendermaßen:

Autonomy refers to systems capable of operating in the real-world environment without any form of external control for extended periods of time. Thus, living

¹¹⁶ *Robotics 3* (1987), Conference Reports, Intelligent Autonomous Systems, 241

¹¹⁷ Ebd., 242

systems are the prototypes of autonomous systems: They can survive in a dynamic environment for extended periods, maintain their internal structures and processes, use the environment to locate and obtain materials for sustenance, and exhibit a variety of behaviors (such as feeding, foraging, and mating). They are also, within limits, capable of adapting to environmental change.¹¹⁸

Zunächst fällt an dieser Definition auf, dass biologische Systeme als Vorbild dienen. Plausibel scheint dies im Hinblick auf die Tatsache, dass die Maschine in der realen Welt, in einem „real-world environment“, eingesetzt wird, ohne dass sie von außen kontrolliert werden kann. Die autonome Aktion eines Systems hat ebenso eine zeitliche Begrenzung. Die Autonomie des Systems hängt von einem größeren Kontext ab, wie dem Ort des Einsatzes, der technologischen Umsetzbarkeit, sowie den Vorgaben und Kontrollen durch den Menschen.

Im Lehrbuch *Künstliche Intelligenz*¹¹⁹ von Lämmel und Cleve wird „autonom“ als eine Eigenschaft von Agenten¹²⁰ wie folgt definiert:

Ein Agent handelt selbstständig bei der Erfüllung der übertragenen Aufgabe. Es werden eigenständige Entscheidungen bezüglich der im Einzelnen auszuführenden Aktionen getroffen.¹²¹

Das ausschlaggebende Kriterium für die Autonomie des sogenannten Agenten ist somit Selbstständigkeit. Daher bleibt zunächst offen, weshalb diese Eigenschaft, die sich auf die Ausführung von Handlungen bezieht, nicht schlichtweg „selbstständig“ sondern „autonom“ bezeichnet wird. Auch dem *Handbuch Robotik* zufolge bedeutet „autonom“ lediglich „unabhängig von äußerer Steuerung“¹²².

Die Autonomie des Agenten besteht darin, dass er stets eine ihm beauftragte Aufgabe erfüllt und somit autonom im Sinne einer eigenen Gesetzgebung nur im Rahmen der ihm übertragenen Aufgabe (d. h. zu deren Lösung) agiert, aber nicht darüber hinaus. Solche „Agenten“ sind im Übrigen als Software- und auch als Hardware-Agenten zu

¹¹⁸ George A. Bekey: *Autonomous Robots. From Biological Inspiration to Implementation and Control*, Cambridge (Massachusetts) 2005, 1

¹¹⁹ Uwe Lämmel / Jürgen Cleve: *Künstliche Intelligenz*, 3. Aufl., München 2008

¹²⁰ Weitere Eigenschaften solcher Agenten sind: „reaktiv“, „zielorientiert“, „kontinuierlich“, „kommunikativ“, „lernend“, „mobil“ und „Persönlichkeit“. Vgl. Lämmel / Cleve 2008, 22. Lämmel und Cleves wortwörtliche Übersetzung des englischen Terminus „agent“ mit „Agent“ im Deutschen ist unglücklich gewählt, zumal der deutsche Begriff sich nicht wie der Englische neutral auf den Handelnden (von lat. agere = handeln) rekurriert sondern vordergründig mit dem Geheimagenten assoziiert wird. Jedoch erklären Lämmel und Cleve ihre Definition damit, dass ein Agent stets Aufgaben im Auftrag einer Person ausführt. Die Art von künstlichen Agenten, von denen hier die Rede ist, wird oft auch mit den Begriffen „intelligent“ oder „mobil“ attribuiert, obgleich keine allgemeinverbindliche Definition existiert. Vgl. ebd. 21.

¹²¹ Lämmel / Cleve 2008, 22

¹²² Haun 2007, 35

verstehen.¹²³ Die Funktionsweise der sogenannten autonomen Agenten gleicht denen der autonomen Systeme.

3.1.2. Gegenstand

Nachdem bereits erläutert wurde, dass Autonomie als Attribut von Agenten und Systemen zu verstehen ist, geht es nun darum, jene autonomen Artefakte – virtuell und physisch – näher zu betrachten, die in der Realität umgesetzt werden. Einer der wichtigsten Bereiche, in denen Hardware-Agenten produziert werden, ist die Robotik, in der mit der Konstruktion und Zusammensetzung sogenannter autonomer Systeme die Interaktion zwischen Informationswissenschaften und physischer Welt gewährleistet werden soll. Bereits 1988 verweist der Herausgeber der Fachzeitschrift *Robotics*¹²⁴ darauf, dass die Robotik sich zunehmend in Richtung autonome Systeme entwickelt:

Starting with this issue our focus will now extend beyond robots, to those systems that contain elements of robots, autonomous systems.¹²⁵

Knasel begreift autonome Systeme als Elemente, die in Robotern, aber auch in anderen Maschinen enthalten sein können. Die technische Autonomie bezeichnet er dabei als „machine autonomy“¹²⁶. Als Anwendungsbereiche nennt er etwa „Autonomous Systems for Space“, und „Autonomous Mobile Robots“ für die industrielle Fertigung.¹²⁷

Der Zielsetzung nach, zu denen autonome Roboter gebaut werden, handelt es sich seither nicht zwingend um autonome Roboter für die Interaktion mit Menschen, sondern vor allem um solche, die für Landerkundung oder militärische Zwecke eingesetzt werden (z. B. „Intelligent Autonomous Functions for Planetary Exploration“¹²⁸, „autonomous underwater vehicles (AUV)“¹²⁹ oder „Crusher Autonomus System“¹³⁰). Das Attribut „autonom“ taucht somit auch in der Definition von Bestandteilen einer bestimmten Art von Agenten oder Systemen auf. Um der Frage nachzugehen, auf welcher Ebene der technologischen Umsetzung in der Robotik das Attribut „autonom“ verwendet wird, ist daher ein Blick in die Fachzeitschrift *IEEE Robotics & Automation*

¹²³ Vgl. Lämmel / Cleve 2008, 22

¹²⁴ *Robotics* 4 (1988)

¹²⁵ T. M. Knasel: Robotics and Autonomous Systems, Editorial in: *Robotics* 4 (1988), 1f

¹²⁶ Ebd. 1

¹²⁷ Vgl. ebd.

¹²⁸ Andreas Birk / Narunas Vaskevicius u.a.: 3-D Perception and Modeling. Motion-Level Teleoperation and Intelligent Autonomous Funktionen, in: *IEEE R&A* 16/4 (2009), 53

¹²⁹ Giacomo Marani / Song K. Choi: Underwater Target Localization. Autonomous Intervention with the DIDSON sonar in SAUVIM, in: *IEEE R&A* 17/1 (2010), 64

¹³⁰ Andrew Bagnell / David Bradley u.a.: Learning for Autonomous Navigation. Advances in Machine Learning for Rough Terrain Mobility. in: *IEEE R&A* 17/2 (2010), 74

*Magazine*¹³¹ aufschlussreich: In ihrem Artikel *Designing Autonomous Robots* definieren Bensalem und Gallien autonome Roboter wie folgt:

Autonomous robots are complex systems that require the interaction of numerous heterogenous software components. Nowadays, robots are getting closer to humans and as such are becoming critical systems that must meet safety properties including logical, temporal, and real-time constraints. [...] Autonomous robots are designed to perform high-level tasks on their own or with very limited external control. They are needed in situations where human control is either infeasible or not cost-effective because of the following features:

- 1) They operate in highly variable, uncertain, and time-changing environments
- 2) They must meet real-time constraints to work properly
- 3) They are often interacting with other agents, both humans and other machines¹³²

Somit entsteht durch das Zusammenspiel verschiedener Softwarekomponenten ein Roboter, der als autonom gilt, sofern er bestimmte Aufgaben nur mit geringer oder ohne Überwachung durch den Menschen erfüllt. Diese Definition scheint auch in anderen Artikeln als Grundlage zu dienen, jedoch ist keine klare definitorische Trennung gegeben zwischen „autonomous robots“¹³³, also dem Roboter als Gesamtsystem, und bestimmten Funktionen und Fähigkeiten des Roboters, welche teilweise auch als autonom bezeichnet werden („autonomous functions“¹³⁴, „autonomous intervention“¹³⁵, „autonomous navigation“¹³⁶). Hier liegt der Verdacht nahe, dass selbst unter den Entwicklern keine klare Ausdifferenzierung vorgenommen wird zwischen (1) dem Code, d.h. der implementierten Software, in der die als autonom bezeichnete Aktionsmöglichkeit programmiert ist, (2) der Hardware, mittels welcher die Aktion physisch sichtbar wird und der (3) Aktion im übertragenen Sinne (ein autonome Aktion als solche, etwa das autonome Navigieren).

Ein reflektierter Umgang mit dem Autonomiebegriff ist zumindest in einem *R&A*-Artikel von Feil-Seifer und Mataric gegeben, in welchem die Autoren zwischen der Autonomie des Menschen und autonomen Robotern wie folgt unterscheiden:

¹³¹ Als Grundlage dienen die Ausgaben *IEEE R&A* 16/1 (2009), 16/3 (2009), 16/4 (2009), 17/1 (2010), 17/2 (2010), 17/4 (2010), 18/1 (2011)

¹³² Saddek Bensalem / Mathieu Gallien u. a.: *Designing Autonomous Robots. Toward a More Dependable Software Architecture*, in: *IEEE R&A* 16/1 (2009), 67. Nähere Erläuterungen zum Aspekt der Kontrolle siehe Unterkapitel 2.1.

¹³³ Siehe Brian R. Cox / Jeffrey L. Krichmar: *Neuromodulation as a Robot controller. A Brain-Inspired Strategy for Controlling Autonomous Robots*, in: *IEEE R&A* 17/1 (2010), 72, sowie Birk / Vaskevicius 2009, 53 und Bensalem / Gallien 2009, 67

¹³⁴ Birk / Vaskevicius 2009, 53

¹³⁵ Marani / Choi 2010, 64

¹³⁶ Bagnell / Bradley 2010, 74

Since autonomy can also refer to robots that are in control of their own actions, we refer to patient/user autonomy as autonomy while referring to the self-control of a robot as robot autonomy or autonomous robots.¹³⁷

Feil-Seifer und Mataric bezeichnen das gesamte System Roboter als autonom, grenzen diese auf Kontrolle reduzierte Art von Autonomie jedoch von der des Menschen ab. Offensichtlich findet ein Nachdenken über den semantischen Gehalt des Autonomiebegriffes erst statt, wenn es sich um die Interaktion von Service-Robotern mit Menschen handelt – einem Bereich also, in dem der Roboter den Menschen nicht für einen mechanischen oder automatisierten Vorgang, sondern für eine Dienstleistung ersetzt und dadurch in die Interaktion mit anderen Menschen treten muss. Hier zeichnen sich bereits erste Probleme ab: Ist der Roboter stets ein Agent, dessen Autonomie sich auf die Erfüllung der ihm zugetragenen Aufgabe beschränkt und damit die Autonomie seiner rechnenden Teilsysteme meint oder wird hier mit dem Begriff Autonomie bereits auf die Sichtbarkeit einer scheinbar autonomen Handlung hingewiesen?

3.2 Ebenen der Mechanisierung und Autonomiegrade

Vom reinen Werkzeug zu jenen als autonom bezeichneten Systemen erfolgt mit der zunehmenden Komplexität der Technologie eine Abnahme der zu Grunde liegenden „Zwecksetzungsautonomie“¹³⁸ des Menschen und ein gradueller Anstieg der Autonomie der Maschine. Nach Gutmann et al. lassen sich technische Systeme in vier Ebenen einteilen, an denen diese graduelle Entwicklung ablesbar ist:¹³⁹

- (1) „Instrumentalisierung“, (z. B. ein Hammer)
- (2) „Maschinisierung“, (z. B. „Autos mit Fahrerinformations- oder Warnsystemen“)
- (3) „Automatisierung“, (z. B. Aufzüge oder automatische Gangschaltung)
- (4) „Autonomisierung“.¹⁴⁰

Während in den ersten beiden Stufen die Zwecksetzung noch vollständig durch den Nutzer erfolgt, d. h. dieser über „Zwecksetzungsautonomie“¹⁴¹ verfügt, nimmt diese ab der Ebene der Automatisierung ab. Der Rückgang der Kontrolle über die Realisierung der Zwecke vollzieht sich bereits ab der Ebene der Maschinisierung, da ab diesem Stadium die Umsetzung der Zwecke den jeweiligen Mitteln übertragen wird. Auf Ebene der Automatisierung wird der Maschine etwa die Fähigkeit der Selbstregulierung durch Vorgabe von Gesetzen eingebaut, während auf Ebene der Autonomisierung das System selbst Strategien zur Selbstregulierung entwickelt. Tim Smithers erklärt dies wie folgt:

¹³⁷ David Feil-Seifer / Maja J. Mataric: Socially Assistive Robotics. Ethical Issues Related to Technology, in: *IEEE R&A* 18/1 (2011), 24-31, 29

¹³⁸ Vgl. Mathias Gutmann / Benjamin Rathgeber / Tareq Syed: Autonome Systeme und evolutionäre Robotik: neues Paradigma oder Missverständnis?, in: Mathias Maring (Hg.): Fallstudien zur Ethik in Wissenschaft, Wirtschaft, Technik und Gesellschaft, Karlsruhe 2011, 185

¹³⁹ Vgl. ebd. 185f

¹⁴⁰ Vgl. ebd., 185f

¹⁴¹ Ebd. 185

[...] automatic systems are self-regulating, but they do not make the laws that their regulatory activities seek to satisfy. These are given to them, or built into them. [...] Autonomous systems on the other hand, are systems that develop, for themselves, the laws and strategies according to which they regulate their behavior: they are self-governing as well as self-regulating.¹⁴²

Smithers beleuchtet zwar treffend den Unterschied zwischen den Ebenen Automatisierung und Autonomisierung. Dennoch ist anzumerken, dass ein System, das den Stand der Autonomisierung erfüllt, in erster Instanz vom Menschen für einen bestimmten Nutzen geschaffen wurde. In der Mensch-Technik-Beziehung ergibt sich in der Ebene der Autonomisierung, z. B. in Form autonomer Systeme, folgende Definition:
143

Autonome Systeme im engeren Sinne sind schließlich solche, bei welchen nicht nur fixierte Funktionen ohne Intervention des Nutzers ausgeführt werden, sondern das System selbst in gegebenen Kontexten sowohl die auszuführenden Funktionen als auch die Formen ihrer Ausführung bestimmt.¹⁴⁴

In einem derartigen System wäre die Rolle des Nutzers beispielsweise auf die Zielsetzung beschränkt und er selbst kann „[...] als Element der Systemumgebung angesehen werden [...]“¹⁴⁵. Dasselbe gilt für autonome Agenten.

Auf Seiten der Technologien ist ein Anstieg der Autonomie festzustellen, je stärker ein autonomes System in den Lebensraum des Menschen greift. Dies zeigt etwa Sebastian Thrun mit der Einteilung verschiedener Roboterarten nach Autonomiegraden.¹⁴⁶ Ihm zufolge lässt sich der gegenwärtige, oder auch der angestrebte Stand der Autonomisierung in folgende drei Bereiche einteilen, die wiederum unterschiedliche, vom Umfeld ihres Einsatzes abhängige Autonomiegrade aufweisen: (1) Bereich der industriellen Nutzung, z.B. Industrieroboter; (2) Professionelle Service-Roboter, z. B. ein „robotic dog“ („professional service robots“); (3) Private Service-Roboter („personal service robots“). Thrun beschreibt die Faktoren der Messbarkeit der Autonomie eines Roboters wie folgt:

[...] the degree of autonomy is often measured by relating the degree at which the environment can be varied to the mean time between failures and other factors indicative of robot performance.¹⁴⁷

Er begründet die Zunahme von Autonomie von Roboter-Art (1) hin zu Roboter-Art (3) mit der zunehmenden Fähigkeit der Roboter – in Abhängigkeit ihrer technischen

¹⁴² Tim Smithers in einem Gespräch mit Luc Steels 1992. Luc Steels: When are robots intelligent autonomous systems?, in: *Robotics and Autonomous Systems* 15 (1995), 5

¹⁴³ Vgl. Gutmann et al. 2011, 185f

¹⁴⁴ Ebd. 186

¹⁴⁵ Ebd. 187

¹⁴⁶ Vgl. Sebastian Thrun: Toward a Framework for Human-Robot Interaction, in: *HUMAN COMPUTER INTERACTION* 19 (2004), 11

¹⁴⁷ Thrun 2004, 14

Leistungs-fähigkeit – in ihrer Umgebung, d.i. die Umgebung des Menschen, Veränderungen vornehmen zu können.¹⁴⁸ Dementsprechend sei gerade in Situationen, in denen Roboter in der Nähe von Menschen arbeiten, wenn nicht sogar mit ihnen zusammen, ein besonders hoher Autonomiegrad erforderlich. Die Forderung nach mehr Autonomie wird mit dem erhöhten Bedarf an Sicherheitsvorkehrungen begründet sowie mit der Tatsache, dass das menschliche Handeln für Roboter weniger vorhersehbar sei als umgekehrt.¹⁴⁹ Folgt man dieser Sichtweise, so wird deutlich, dass der Mensch den nicht-determinierten Faktor in einer solchen Mensch-Maschine-Interaktion darstellt und die „Autonomie des Roboters“ ein dem Roboter integrierter Anpassungsmechanismus an den Menschen und dessen Umwelt darstellt. Roboter und autonome Systeme dieser Art haben, trotz ihres hohen Autonomiegrades, einen vom Menschen gesetzten Zweck, zu dessen Erreichung eine gewisse Form von Selbstständigkeit erforderlich ist. Für diese Form schlage ich die Bezeichnung „offene autonome Systeme“ vor.¹⁵⁰

Diesen stehen „geschlossene autonome Systeme“ gegenüber, für die in erster Linie zwar obige Autonomisierungsdefinition gilt. Der Unterschied jedoch liegt in der Geschlossenheit des Systems, in das der Mensch keine zielorientierten Vorgaben mehr machen könnte. Das System leistete in einem solchem Fall selbst „nicht nur die *Kontrolle*, sondern auch die *Erstellung* des Systems“¹⁵¹.

3.3. Kriterien technischer Autonomie

Bislang hat sich die Definition ergeben, dass Autonomie als ein Attribut benutzt wird, das auf ein gewisses dem System implementiertes technisches Level rekurriert, welches das System in einer vorgegebenen Zeit und einem definierten Kontext einsetzt. Ein wesentlicher Faktor der technischen Autonomie ist zum einen Kontrolle, die, je nach Grad der technologischen Stands, eine zunehmende Selbstkontrolle der Maschine und eine abnehmende Kontrolle durch den Menschen bedeutet. Der Faktor künstliche Intelligenz ist für die Autonomie von Technologien in einer besonderen Weise ausschlaggebend: Der Anspruch, autonome Systeme zu bauen, ist überhaupt erst in der jüngeren KI-Forschung entstanden. Technische Autonomie ist insofern ohne KI nicht denkbar, umgekehrt besteht jedoch kein Abhängigkeitsverhältnis der KI von Autonomie. Der Einfluss von Kontrolle und KI auf technische Autonomie soll daher in den folgenden Unterkapiteln näher untersucht werden.

3.3.1. Kontrolle

Unabhängig vom jeweiligen Autonomiegrad ist allen Beispielen aus dem vorherigen Kapitel gemein, dass Autonomie stark an den Begriff der Kontrolle gebunden ist. Dies ist zum einen die (1) Kontrolle von außen und zum anderen (2) Kontrolle von innen.

¹⁴⁸ Vgl. Thrun 2004, 14f

¹⁴⁹ Vgl. ebd., 15

¹⁵⁰ Siehe hierzu Kapitel 3.3.1.

¹⁵¹ Gutmann et al. 2011, 187

(1) Äußere Kontrolle durch den Menschen als „supervisor“ ist zunächst jene Form, wie sie aus der Vorstufe Maschinisierung bekannt ist. Je fortgeschrittener der Grad der technischen Entwicklung, umso geringer der Grad der Kontrollierbarkeit. Ziel solcher Entwicklungen ist die Vereinfachung der Bedienung für den Menschen; Cheng und Zelinsky fassen dies unter das Paradigma der „supervised autonomy“¹⁵²:

The general concept of our paradigm is to incorporate supervisory control with a qualitative approach for the control of robots. [...] The approach we have taken shifts all basic autonomous functions to the physical robot agent, integrated with a set of qualitative instructions, in combination with a simple graphical user interface, and together with suitable feedback form the complete framework.¹⁵³

Nach diesem Paradigma ist eine Rückkopplung sämtlicher Aktionen des Roboters an den Nutzer vorhanden. Das „user interface“ ist nicht nur als Kontrollmittel, sondern auch als Schnittstelle gedacht, über die der Roboter Informationen erhält. Cheng und Zelinsky bezeichnen den Nutzer als Teil des „framework“, welches das Roboter-System zum Funktionieren benötigt. Somit handelt es sich hierbei um offene autonome Systeme.

(2) Eine neues Level von Kontrolle soll mittels der dem Roboter implementierten Selbstkontrolle erreicht werden. Diese überwacht die autonomen, nicht-determinierten Aktionen und wäre damit der bereits erläuterten „human supervisory“ nicht mehr bedürftig. Hierfür wird beispielsweise die Neuromodulation erforscht: Die Kontrolle der Autonomie des Roboters soll durch strukturelle Ähnlichkeit im Aufbau zum menschlichen Gehirn erreicht werden.¹⁵⁴ Die Selbstkontrolle des Roboters ist auf fehlerfreie Ausführung von Aufgaben gerichtet, wobei der Roboter seine Umwelt und die Ergebnisse seiner Aktionen auswerten und daraus „lernen“ soll:

We have shown how a model of the neuromodulatory system and surrounding regions can cause a robot to sharpen its sensory systems, attend to behaviorally relevant objects and ignore distractions, learn to predict the value and outcome of its decisions, and respond decisively and appropriately to environmental events.¹⁵⁵

Selbst wenn in einem solchen Fall eine Simulation menschlicher Kontrollfunktionen entwickelt würde, die sich auf die Autonomie beziehen soll, so bliebe dennoch festzuhalten, dass es sich um das kontrollierte Ausführen von Aktivitäten handelt, die im Falle des autonomen Roboters vom System selbst und nicht vom Menschen programmiert wurden. Jedoch unterliegen diese Aktivitäten weiterhin der Zielsetzung und Funktion des Roboters als spezialisiertes, sich selbst regulierendes Werkzeug. Um dies in einem Beispiel auszudrücken: Ein Industrieroboter würde seine eigenen Funktionen

¹⁵² Gordon Cheng / Alexander Zelinsky: Supervised Autonomy: A Framework for Human-Robot Systems Development, in: *Autonomous Robots* 10 (2001), 252

¹⁵³ Cheng / Zelinsky 2001, 252

¹⁵⁴ Vgl. Cox / Krichmar 2010, 73

¹⁵⁵ Cox / Krichmar 2010, 73

als Industrieroboter optimieren (die der Mensch dann unter Umständen nicht mehr nachvollziehen kann) und sich nicht beispielsweise in einen Serviceroboter umprogrammieren – zumal hierzu auch die materielle Beschaffenheit als Hindernis zu berücksichtigen ist.

Autonomie und Kontrolle liegen also im Detail und der erwartete Autonomiegrad der Fähigkeiten und Funktionen, die von den jeweiligen Systemen umgesetzt werden sollen, sind nicht einheitlich. Daher liefern Bertschinger et al. einen ersten Vorschlag, auf informationstheoretischer Grundlage den Autonomiegrad in mathematischen Formeln auszudrücken.¹⁵⁶ Wesentlich ist für die Messbarkeit von Autonomie der Umfang, in welchem das System in seine Umgebung verändernd eingreifen und Kontrolle über diese ausüben kann.

3.3.2. Intelligenz

Beginnend mit einer Beschreibung des Intelligenzbegriffes an sich, sollen in diesem Kapitel vornehmlich die verschiedenen Herangehensweisen der KI-Forschung beschrieben werden, die als Grundlage für die Entwicklung autonomer Systeme dienen.

3.3.2.1. Herkunft des Intelligenzbegriffes

Der Begriff „Intelligenz“ (lat.: *intellegentia*) leitet sich vom lateinischen Verb *intellegere* (gebildet aus lat. *inter* = zwischen und *legere* = (aus-) wählen, lesen, sammeln) ab, welches bereits in der Antike den geistigen Prozess des „dazwischen Auswählens“, des Verstehens und Erkennens durch kritische Auswahl meint.¹⁵⁷

Bereits Cicero gebraucht den Begriff *intellegentia* als Synonym für *ratio*, das „höhere Seelenvermögen“¹⁵⁸ und die Fähigkeit zu erkennen (*intellectus*).¹⁵⁹ Mit großem Interesse widmen sich insbesondere Philosophen des Mittelalters und der Scholastik dem Begriff Intelligenz. Er dient zur Beschreibung des „Geistwesens Gottes“¹⁶⁰ und himmlischer Wesen, sowie der menschlichen Fähigkeit, diese zu erkennen.¹⁶¹ Thomas von Aquin hält folgende vier Bedeutungen des Begriffes Intelligenz fest:

- a) vernünftige Substanz, b) Vernunfttätigkeit, Vernunftkenntnis, c) Vernunfteinsicht, d.i. unmittelbare Vernunftkenntnis, d) Verständnis, intellektuelle Auffassung (synonym mit <intellectus> [...])¹⁶²

¹⁵⁶ Siehe Nils Bertschinger et al.: Autonomy: An information theoretic perspective, in *BioSystems* 91 (2008), 331-345

¹⁵⁷ Vgl. Meyers Enzyklopädisches Lexikon: Artikel Intelligenz, Bd. 12, 9. Aufl., Mannheim 1974

¹⁵⁸ Rainer Piepmeier: Intelligenz, *Intelligensia, Intellektueller*, in: HWPh Bd. 4 (I-K) 1976, 446

¹⁵⁹ Vgl. ebd.

¹⁶⁰ Ebd.

¹⁶¹ Vgl. ebd. Auch Hugo von St. Victor bezeichnete Intelligenz als „das höchste Seelenvermögen zur Schau Gottes, zur *contemplatio dei*.“

¹⁶² Piepmeier 1976, 446

Aus dieser Definition heraus etabliert sich die *intelligentia divina* zu einem ausschlaggebenden Charakteristikum Gottes, was sich wiederum im Französischen (*intelligence suprême*) sowie dem Englischen (*supreme intelligence*) als Bezeichnung für Gott niederschlägt. Im Hinblick auf die folgenden Ausführungen zur künstlichen Intelligenz halte ich auch die scholastische Vorstellung intelligenter himmlischer Wesen für erwähnenswert: Thomas von Aquin verweist auf arabische Quellen, in denen Engel – materielle Substanzen (*substantiae seperatae*) – als Intelligenzen bezeichnet werden, da diese „immer aktuell erkennen [sie können nicht schlafen].“¹⁶³ Vor diesem Hintergrund gewinnt man dem *Ubiquitous Computing*¹⁶⁴ eine neue Interpretation ab, denn die Vorstellung von der Allgegenwart immer denkender, nicht schlafender Intelligenzen – den Engeln also – ähnelt stark der Idee des *Ubiquitous Computing*, mittels dessen die rechnergestützte Informationsverarbeitung der den Menschen umgebenden Technologien zwar allgegenwärtig, aber dennoch unsichtbar sein soll. Im Übrigen hält sich bis zur kopernikanischen Wende die christliche Vorstellung einer himmlischen Intelligenz, die in Anlehnung an die aristotelischen körperlosen Intelligenzen, aus hierarchisch angeordneten Engeln besteht, die „als Beweger der Sphären des Himmelsgewölbes gelten.“¹⁶⁵

Neben Vernunft- und Erkenntnisfähigkeit wird der Bedeutungsgehalt von Intelligenz um die Komponenten des Göttlichen und der systematischen Struktur eines Kollektivs erweitert. Hiervon ausgehend wird beispielsweise im französischen Absolutismus Intelligenz, die nicht auf das Erkenntnisvermögen des einzelnen beschränkt ist sondern einer systematischen, göttlichen Überordnung gleicht, zur Bezeichnung des Monarchen und seiner Minister (*Intelligence de l'Etat*).¹⁶⁶ In den Sozialwissenschaften hingegen werden unter anderem jene gesellschaftlichen Gruppen als ‚Intelligenz‘ bezeichnet, die für „historische Resultate der Herausbildung der bürgerlichen Gesellschaft“ verantwortlich waren, wie etwa freie Schriftsteller und Philosophen.¹⁶⁷ Innerhalb jener als Intelligenz bezeichneten Gruppe, etwa im Idealismus, fand wiederum eine gezielte Erörterung des Intelligenzbegriffes aus philosophischer Sicht statt. So etablierte sich im

¹⁶³ Piepmeier 1976, 446. In Thomas von Aquins *Summa Theologiae I* lautet die Passage wie folgt: „In quibusdam tamen libris de arabico translatis substantiae seperatae, quas nos Angelos dicimus Intelligentiae vocantur, forte propter hoc quod uisusmodi substantiae semper actu intelligunt.“ (Zit. nach Piepmeier 1976, 446)

¹⁶⁴ Siehe Mark Weiser: The Computer for the 21st Century, in: *Scientific American*, Sept. 1991, 94-104. In diesem Artikel führt Weiser die Bezeichnung „*Ubiquitous Computing*“ für die computergestützte Steuerung von Gegenständen ein. Weiser gehörte 1991 zu den ersten Wissenschaftlern, die an der Verwirklichung des *Ubiquitous Computing* arbeiteten (am Xerox Palo Alto Research Center).

¹⁶⁵ Piepmeier 1976, 446

¹⁶⁶ Ebd. 447

¹⁶⁷ Sozialwissenschaftliche Definitionen von Intelligenz orientieren sich stets an „sozialen Einstellungen oder Handlungstypen, [...] Funktionen des kulturellen oder des sozialen Systems.“ (Röttgers, Kurt: Intelligenz; Intellektueller II. V (soz.), in: HWPh Bd. 4 (I-K) 1976, 448).

deutschen Idealismus ein philosophisch-psychologischer Intelligenzbegriff als typisch menschliches Charakteristikum.

3.3.2.2. Definition Intelligenz heute

Vor diesem Hintergrund hat sich das allgemeine, psychologisch geprägte Verständnis herausgebildet, nach welchem Intelligenz die Fähigkeit bezeichnet, abstrakte Beziehungen herzustellen und zu erfassen sowie neue Situationen durch problemlösendes Verhalten zu bewältigen.¹⁶⁸ Kennzeichnend für das gegenwärtige Verständnis menschlicher Intelligenz ist ihre Messbarkeit durch Intelligenztests, wie sie seit Beginn des 20. Jahrhunderts in der wissenschaftlichen Psychologie praktiziert wird. Demnach bezeichnet der Intelligenzquotient das „[...] Maß für die allgemeine intellektuelle Leistungsfähigkeit, die sich aus dem Verhältnis von Intelligenzalter (IA) zum Lebensalter (LA) nach der Formel $IQ = (IA/LA) \cdot 100$ ergibt.“¹⁶⁹ Anhand solcher Tests ist beispielsweise ein Verlauf der Entwicklung der Intelligenz ablesbar.

Nach einer Periode starker positiver Beschleunigung in der frühen und mittleren Kindheit verlangsamt sich die I-entwicklung ab dem 10. Lebensjahr bis zum Erreichen des Erwachsenenalters. Zwischen dem 20. und 30. Lebensjahr setzt allmählich ein Verfall der I-leistungen ein.¹⁷⁰

Intelligenz setzt sich jedoch aus verschiedenen Faktoren wie sprachliches Verständnis, Assoziationsflüssigkeit, Rechengewandtheit, Gedächtnis, räumliches und schlussfolgerndes Denken sowie Auffassungsgeschwindigkeit zusammen.¹⁷¹ Wesentlich beeinflusst wird die Entwicklung der Intelligenz durch das Zusammenwirken von genetischer Veranlagung und Umweltbedingungen. Letzteres umfasst insbesondere das kulturelle und soziale Milieu, in dem sich der Einzelne bewegt, weshalb aus psychologischer Sicht die Entwicklung der Intelligenz trotz gleicher Erbanlagen unter der Gegebenheit kultureller und sozialer Unterschiede abweichend sein kann.¹⁷² Auffällig ist, dass auf die Phase der Messbarmachung menschlicher Intelligenz in der Psychologie unmittelbar das Zeitalter der intensiven Forschung zur künstlichen Intelligenz folgt. Der in der breiten Öffentlichkeit etablierte, von KI-Forschern forcierte Trugschluss, dass Intelligenz, da künstlich exakt messbar, auch künstlich rekonstruierbar sei, scheint aus historischer Distanz nachvollziehbar.

¹⁶⁸ MEL, Artikel Intelligenz

¹⁶⁹ Ebd. Ein Grad von 100 bildet den Durchschnitt. Erstmals vorgeschlagen wurde der Intelligenzquotient von W. Stern (1912), welcher ihn auf der Basis verschiedener Tests elaborierte, die an Kindern durchgeführt wurden. Vgl. ebd.

¹⁷⁰ MEL, Artikel Intelligenz

¹⁷¹ MEL, Artikel Intelligenz

¹⁷² Nachgewiesen wurde dies insbesondere durch die Zwillingsforschung. Vgl. MEL

3.3.2.3. Künstliche Intelligenz

Die künstliche Intelligenz, Teilbereich der Computerwissenschaften, befasst sich mit der Entwicklung von Programmen, die aufgrund ihres Outputs¹⁷³ als intelligent bezeichnet werden können. Als Parameter von Intelligenz gelten dabei Fähigkeiten des Menschen, welche nun als „informationsverarbeitende Prozesse [betrachtet werden, die] naturwissenschaftlichen Untersuchungsmethoden (und ingenieurmäßiger Verwendung) zugänglich“¹⁷⁴ sind. Das Verhalten des Menschen und seine Fähigkeiten sind jedoch auf unterschiedliche Art und Weise definierbar, wodurch sich im Laufe der Zeit zwangsläufig verschiedene Herangehensweisen an die künstliche Intelligenz ergeben haben, welche dementsprechend jeweils eigenen Definitionen folgen. Stuart J. Russel und Peter Norvig teilen in der dritten Auflage ihres Buches *Artificial Intelligence*¹⁷⁵ diese verschiedenen Herangehensweisen grob in vier Bereiche: (1) *acting humanly*, (2) *thinking humanly*, (3) *thinking rationally* sowie (4) *acting rationally*.¹⁷⁶ Jene Methoden, die das menschliche Handeln und Denken zum Vorbild nehmen, werden von einem empirischen Ansatz angegangen, d. h. die Beobachtung und Evaluation menschlicher Verhaltensmuster dient der technischen Nachbildung als Grundlage. „Rationale“ Herangehensweisen sind jene, die aus der Mathematik und den Ingenieurwissenschaften heraus entstanden sind.¹⁷⁷

(1) „Acting Humanly: The Turing Test approach“

Der Ansatz *acting humanly* ist der älteste, der in der KI-Forschung verfolgt wird. Russel und Norvig sprechen in diesem Fall auch von „The Turing Test Approach“, da die Grundidee des *acting humanly*-Ansatzes in der Konstruktion von Maschinen liegt, welche den „Turing-Test“¹⁷⁸ bestehen können. Alan Turing, Namensgeber jenes Tests,

¹⁷³ Ich vermeide an dieser Stelle bewusst die anthropologisch bzw. biologisch präformierten Termini „Verhalten“ oder „Handlung“. Siehe hierzu Kap. 5.1.

¹⁷⁴ Jörg Siekmann: Künstliche Intelligenz: von den Anfängen in die Zukunft, in: Günther Cyranek / Wolfgang Coy (Hgg.): Die maschinelle Kunst des Denkens, Perspektiven und Grenzen der künstlichen Intelligenz, Braunschweig / Wiesbaden 1994, 11

¹⁷⁵ Stuart J. Russell / Peter Norvig: *Artificial Intelligence. A Modern Approach*, Boston / München u. a. 2010

¹⁷⁶ Vgl. ebd., 2

¹⁷⁷ Ebd., 1f

¹⁷⁸ Siehe Alan M. Turing: Computing, Machinery and Intelligence, in: *Mind* 59/236 (1950), 433-460. Der Test ist im Grunde ein Fragespiel, in dem der Computer von seinem menschlichen Gegenüber aufgrund der Antworten, welche er auf bestimmte Fragen gibt, für einen Menschen gehalten wird. Turing ersetzt die Frage „Can machines think?“ durch das „imitation game“ (Imitationsspiel), wodurch er von vorn herein nicht nachzuweisen versucht, dass Maschinen denken können, sondern lediglich, dass sie von Menschen unter Erfüllung bestimmter Kriterien aufgrund ihres Verhaltens für intelligent gehalten werden könnten. Einen relativ aktuellen, kompakten Überblick und eine Bewertung der Entwicklung des jahrzehntelangen Turing-Test-Diskurses bietet folgendes Paper: Ayse Pinar Saygin / Ilyas Cicekli / Varol Akman: Turing Test: 50 Years Later, in: *Minds and Machines* 10 (2000), 463-518.

verstand unter „denkenden Maschinen“ elektronische oder digitale Computer,¹⁷⁹ deren Intelligenz aufgrund folgender Fähigkeiten nachgewiesen werden kann: (1) „natural language processing“¹⁸⁰ (Kommunikation in einer Sprache), (2) „knowledge representation“¹⁸¹ (Möglichkeit zur Wissensspeicherung), (3) „automated reasoning“¹⁸² (Nutzung der gespeicherten Informationen, um Antworten zu generieren sowie neue Schlüsse zu ziehen), (4) „machine learning“¹⁸³ (Anpassung an neue Gegebenheiten, Extrapolation neuer Verhaltensmuster).

Die wesentliche Parallelität von menschlichen und elektronischen Maschinen liegt Turing zufolge in der Fähigkeit, mathematische Funktionen zu verstehen, Daten zu speichern und sie zu verarbeiten.¹⁸⁴ Das eigentliche Anliegen der Forschung nach diesem Muster ist das Verständnis und die Entschlüsselung jener Prinzipien, die der Intelligenz zugrunde liegen.¹⁸⁵ Es wird davon ausgegangen, dass künstliche Intelligenz dann vorhanden ist, wenn eine Maschine einen menschlichen Denkprozess simulieren kann. Turing reduzierte das Funktionieren einer quasi denkenden Maschine bereits selbst auf die Simulation menschlicher Verhaltensweisen. Das heißt, eine Maschine verhält sich in einer bestimmten Situation so, als liefere sie das Ergebnis eines Denkprozesses, wie man ihn von einem Menschen erwarten würde. In Wirklichkeit ist die Programmstruktur, die bei der Maschine Input und Output steuert, jedoch nicht mit dem Denkprozess eines Menschen vergleichbar. Künstliche Intelligenz umfasst somit nur einen kleinen Teilbereich dessen, was die menschliche Intelligenz hervorbringen kann. Um künstliche Intelligenz nachzuweisen ist daher eine physische Simulation des Menschen nicht erforderlich.

Der Entwicklung der KI-Forschung der letzten 60 Jahre und ihren gegenwärtigen Ambitionen nach zu urteilen, müsste Russel und Norvig zufolge der Turing-Test jedoch dahingehend vervollständigt werden, dass für den Nachweis von KI im Sinne des „acting humanly“-Ansatzes auch die Fähigkeiten (5) „computer vision“ und (6) „robotics“ notwendig wären.¹⁸⁶ Maschinen, die nach jenen sechs Fähigkeiten entwickelt

¹⁷⁹ Vgl. Turing 1950, 436. Turing führt in 435f an, dass elektronische bzw. digitale Computer die einzigen sind, die für die technische Entwicklung und die Prüfung durch das Imitationsspiel zweckmäßig erscheinen.

¹⁸⁰ Russell / Norvig 2010, 2

¹⁸¹ Ebd.

¹⁸² Ebd.

¹⁸³ Ebd.

¹⁸⁴ Vgl. Turing 1950, 439: „In certain computers the storage system is mainly acoustic. The feature of using electricity is thus seen to be only a very superficial similarity. If we wish to find such similarities we should rather for mathematical analogies of function.” Turing ist mit seinem Essay *Computing, Machinery and Intelligence* somit auch als einer der ersten Vertreter der „thinking rationally approach“ zu betrachten.

¹⁸⁵ Vgl. Russell / Norvig 2010, 3

¹⁸⁶ Russell / Norvig 2010, 3

werden, dienen heutzutage jedoch insbesondere dem Einsatz im Servicebereich sowie in der Industrie oder im militärischen Bereich. Der Zweck der Forschung ist anwendungsbezogen und besteht nicht allein darin, eine Maschine zu bauen, die den Turing-Test besteht – oder wie es Russel und Norvig an einem Beispiel verdeutlichen:

Aeronautical engineering texts do not define the goal of their field as making ‚machines that fly so exactly like pigeons that they can fool even other pigeons‘.¹⁸⁷

(2) „Thinking humanly: The cognitive modeling approach“¹⁸⁸

Unter *thinking humanly* werden jene Methoden subsumiert, deren Entwicklung eng mit den Kognitionswissenschaften zusammenhängen. KI-Forschung als Teilbereich der Kognitionswissenschaften soll dazu beitragen, das menschliche Denken zu entschlüsseln. Der *thinking humanly*-Ansatz ließe sich folgendermaßen verallgemeinern:

Once we have a sufficiently precise theory of the mind, it becomes possible to express the theory as a computer program. If the program’s input-output behavior matches corresponding human behavior, that is evidence that some of the program’s mechanisms could also be operating in humans.¹⁸⁹

Hierzu muss jedoch zunächst die Funktionsweise des menschlichen Denkens entschlüsselt werden, etwa durch Introspektion oder durch psychologische Experimente. Die Kognitionswissenschaften sind daher ein interdisziplinäres Gebiet, in dem experimentelle Versuche, etwa aus der Psychologie und Neurophysiologie, als Computermodelle dargestellt und somit als ein theoretisches Abbild des menschlichen Denkens angesehen werden.¹⁹⁰ Als Grundlage dienen stets Experimente an Menschen oder Tieren. Russel und Norvig weisen jedoch darauf hin, dass der Anspruch menschliches Denken modellhaft nachzubilden auf zwei verschiedene Weisen verstanden werden kann: (1) Der interdisziplinäre Ansatz aus den Kognitionswissenschaften versucht, durch gleiche Strukturen in künstlicher und natürlicher Intelligenz, Schnittstellen zwischen Mensch und Technik zu verwirklichen. (2) Die Modellhaftigkeit des menschlichen Denkens kann beispielsweise auch ein Algorithmus sein, der zu einer mathematischen Aufgabe dieselbe Lösung berechnet wie der Mensch.¹⁹¹

(3) „Thinking rationally: The ‘laws of thought’ approach“¹⁹²

Die im vorherigen Abschnitt erwähnte zweite Deutungsmöglichkeit von „menschlich denkenden“ Programmen ist teilweise besser mit dem *thinking rationally*-Ansatz zu erklären. Dieser Ansatz der KI-Forschung ist aus der klassischen Logik hervor-

¹⁸⁷ Russell / Norvig 2010, 3

¹⁸⁸ Ebd.

¹⁸⁹ Ebd.

¹⁹⁰ Vgl. ebd.

¹⁹¹ Russell / Norvig 2010, 3

¹⁹² Ebd., 4

gegangen. Künstliche Intelligenz bezieht sich in diesem Fall auf problemlösungsorientierte Programme, die mit Prädikatenlogik auskommen. Ein wesentlicher Nachteil dieser Methode ist, dass sie nur mit hundertprozentig sicherem Wissen nützlich ist.¹⁹³

(4) „Acting rationally: The rational agent approach“¹⁹⁴

Der *acting rationally*-Ansatz ist als Erweiterung und Kombination der bereits vorgestellten Herangehensweisen zu betrachten und vereint die heutigen Forschungsziele der KI. Künstliche Intelligenz soll mit diesem Ansatz in *Rational Agents* umgesetzt werden, wodurch eine Art Turing-Maschine entsteht. Die Anforderungen an solche *Rational Agents* gehen über die eines Computerprogramms hinaus:

[...][They] operate autonomously, percieve their environment, persist over a prolonged time period, adapt to change, and create and pursue goals¹⁹⁵

Rational Agents zeichnen sich dadurch aus, dass ihre Aktionen nicht auf richtige oder falsche Inferenzen basieren, sondern darauf abzielen, ein bestmöglich zu erwartendes Output zu liefern. Der *thinking rationally*-Ansatz dient zwar als Grundlage, doch sind rational richtige Folgerungen hierbei nur eine von vielen Möglichkeiten zur Aufgabenlösung. Mit diesem Ansatz soll die Bewältigung von Aufgaben geleistet werden, die nicht bereits vorprogrammiert sind, sondern die sich den *Rational Agents* erst in der Konfrontation mit der Umwelt stellen. Die Rahmenbedingungen zur Konstruktion einer Turing-Maschine sind viel zu kompliziert und vielfältig, als dass sie alle bereits vorprogrammiert werden könnten. Daher liegt das Innovative des *acting rationally*-Ansatzes darin, Maschinen derart zu konstruieren, dass ihre Aktionen nach ihrem bestmöglichem Gelingen bewertet werden. In diesem Ansatz kommt nun Autonomie in einer besonderen Weise zum Tragen: Sie bezeichnet die Unabhängigkeit des *Rational Agents* von seinem inputgebenden Designer:

A rational agent should be autonomous – it should learn what it can to compensate for partial or incorrect prior knowledge. [...] So, just as evolution provides animals with enough built-in reflexes to survive long enough to learn for themselves, it would be reasonable to provide an artificial intelligent agent with some initial knowledge as well as an ability to learn.¹⁹⁶

Die Autonomie des *Rational Agent* spielt sich stets innerhalb einer „task environment“¹⁹⁷, also dem gegebenen Umfeld, zu dessen Zweck der *Rational Agent* eingesetzt wird, ab. Sie bezeichnet die Möglichkeit, trotz unsicherem Wissen aufgrund von „Gelerntem“, d. h. aus der Umwelt generiertem Wissen, eine Entscheidung zu

¹⁹³ Vgl. Russell / Norvig 2010, 4

¹⁹⁴ Russell / Norvig 2010, 4

¹⁹⁵ Ebd.

¹⁹⁶ Ebd., 39

¹⁹⁷ Ebd., 38

treffen und eine Aktion auszuführen. Die Grundidee hinter dieser Vorgehensweise ist nach Russell und Norvig die Lernfähigkeit von Tieren und Menschen.¹⁹⁸

Umfeld und Umwelt sind dabei stets dasjenige, in das die Maschine vom Menschen mit einem bestimmten Ziel hineingesetzt wurde. Technische Autonomie verweist lediglich darauf, dass der Mensch der Maschine mehr Eigenständigkeit in der Lösung von Problemen überträgt, indem ihr mehr Möglichkeiten zur Selbstkontrolle und Aufgabenlösung aufgrund von eigenem Generieren äußerer Rahmenbedingungen vorgegeben werden. Somit definieren Russell und Norvig Autonomie als die Nutzungsart von Wissen und Lernen für die Aufgabenbewältigung eines *Rational Agent*; autonom ist nicht lediglich ein Programm, sondern das Artefakt, das aktiv verändernd in seine Umwelt, d. i. die Umwelt des Menschen, eingreift.

3.4. Zusammenführung und Fortgang

Ich habe in diesem Kapitel aufgezeigt, dass Autonomie in den neuen Technologien einen vollkommen anderen Bedeutungsgehalt besitzt, als die Definition der Autonomie des Menschen. Beiden gleich ist lediglich der rein begriffliche Bezug auf die Selbstgesetzgebung, die im technischen Bereich jedoch auf die technische Lösung einer Aufgabenstellung abzielt.

Der Kategorisierung der Methoden der KI-Forschung von Russell und Norvig ist zu entnehmen, dass die technische Entwicklung stark vom jeweiligen zugrundeliegenden Menschenbild abhängt, was dementsprechend zu weit gestreuten Vorstellungen über die Potentiale von KI führt. So zeigt etwa die langjährige Debatte um starke und schwache KI¹⁹⁹, dass es durchaus überzeugte Meinungen gibt, künstliche und natürliche Intelligenz als äquivalent zu betrachten. Es ist also naheliegend, dass durch den jeweiligen Ansatz in der KI auch unterschiedliche Vorstellungen darüber entstehen, wie bestimmte Eigenschaften wie Autonomie – und damit einhergehend Kontrolle – umgesetzt und definiert werden sollen.

Es bleibt daher im Fortgang dieser Arbeit den Zweck des Gebrauchs technischer Mittel zu konkretisieren. Vor dem Hintergrund der zunehmenden Vernetzung von Mensch und Technik resultieren insbesondere folgende zwei Problemfelder:

(1) Zum einen begegnen wir einem realen Anstieg an Schnittstellen zwischen Mensch und Technik, in denen Autonomie und technische Autonomie aufeinander treffen. Diese zunehmende Komplexität technischer Artefakte und ihre Interaktion mit dem Menschen wirft die Frage auf, welche Ausprägung von Autonomie als menschlich, und welche als technisch betrachtet werden kann, bzw. wann die menschliche Autonomie aufhört und die technische anfängt. Dies ist zum Beispiel abhängig von der Reichweite der *human*

¹⁹⁸ Zur Übertragung von Begriffen der Anthropologie in die Biologie und schließlich in die Robotik siehe Kapitel 5.1.1.

¹⁹⁹ Siehe Fn. 2.

*supervisory*²⁰⁰ innerhalb offener autonomer Systeme. Daher ist in Kapitel 4 zunächst herauszuarbeiten, nach welchen Kriterien Autonomie auch heute noch als menschen-spezifisches Charakteristikum gilt, d.h. welches jene Fähigkeiten sind, die einem autonomen Menschen, aber nicht einer autonomen Maschine zukommen.

(2) Zum anderen sind zunehmende sprachliche Schwierigkeiten im Verständnis technischer Systeme zu verzeichnen, was von den unzureichenden Termini innerhalb der Technik sowie darüber hinaus herrührt. Es stellt sich die Frage, inwieweit unzureichende Begrifflichkeiten auch heute noch zu einem falschen Verständnis der Leistungsfähigkeit, der Unterschätzung oder auch Überschätzung einer Technik führt.²⁰¹ Da jene Maschinen vom Menschen selbst geschaffen und als autonom bezeichnet werden, bleibt noch zu untersuchen, wie viel Autonomie wir den Maschinen übertragen wollen, bzw. zu welchem Zweck sie autonom sein müssen.

²⁰⁰ Cheng / Zelinsky 2001

²⁰¹ Ein klassisches Beispiel hierfür ist immernoch Joseph Weizenbaums Computerprogramm *Eliza* aus dem Jahr 1966. Weizenbaums Anliegen war es, das Sprachverstehen eines Computerprogramms zu demonstrieren. Testpersonen, zum Teil sogar Psychiater, sahen hierin jedoch missverständlicherweise ein Tool der automatisierten Psychotherapie. Siehe hierzu Joseph Weizenbaum: *Die Macht der Computer und die Ohnmacht der Vernunft*, aus dem amerikan. übers. von Udo Rennert (*Computer Power and Human Reason. From Judgement to Calculation*, 1976), Frankfurt a. M. 1977; Joseph Weizenbaum: *Eliza – a computer program for the study of natural language communication between man and machine*, in: *Communications of the ACM* 9/1 (1966), 36-45; Jens Schanze: *Plug & Pray*, Dokumentarfilm, Mascha Film Judith Malek-Mahdavi & Jens Schanze GbR 2009

Do Machines Think about Machines Thinking?

(Leonard Pinsky, 1951, Titel des ersten Leserbriefs zu Turings Essay Computing,
Machinery and Intelligence)

4. Autonomie kommt nur dem Menschen zu

Nachdem ich in den vorangegangenen Kapiteln bereits den Autonomiebegriff in der Philosophie sowie als Attribut von Technologien erläutert habe, soll nun erörtert werden, nach welchen Kriterien Autonomie eine Eigenschaft ist, die lediglich dem Menschen, nicht aber der Maschine zukommt.

Zwar gibt es allgemeinverbindliche Definitionen dessen, was wir konkret unter Maschinen und neuen Technologien verstehen – nicht aber, was der Mensch ist. Die Bestimmung der Eigenschaften des Menschen hängt wesentlich vom Kontext ab und kann biologisch, theologisch, naturwissenschaftlich oder philosophisch begründet werden. Somit ist die Antwort auf die Frage was der Mensch ist, in einem semantisch unscharfen Feld verankert, denn abgesehen von der klaren biologischen Definition als *homo sapiens sapiens*, gelten für den Menschen gleichzeitig viele andere Beschreibungen, etwa als *zoon logon echon*, *zoon politikon*, *homo rationale*, *homo ridens*, oder *homo ludens*.²⁰² Der Mensch besitzt als „flexibles Vielfachwesen“²⁰³ die Fähigkeit, sich selbst als autonomes und handelndes Wesen zu definieren. Autonomie ist, wie in Kapitel 2 gezeigt, eine Fähigkeit, die der Mensch im Laufe der Zeit für sich selbst entdeckt und sich selbst zugesprochen hat. Wenn also einem technischen Artefakt Autonomie abgesprochen werden soll, dann erfolgt dies über die Negation bestimmter Fähigkeiten, die gerade aufgrund seiner technischen Beschaffenheit nicht realisierbar sind. Diese Fähigkeiten oder Faktoren sind etwa Freiheit, Personalität, Moralität, Bewusstsein oder Sprachverstehen.

In Kapitel 4.1 geht es zunächst um die Frage, wieso der Mensch sich selbst als autonom handelndes Wesen begreift, d.h. autonom in einer anderen Art als es die Maschine sein soll. Ich werde mich dabei einerseits auf die moralische und andererseits auf die pragmatische Motivation bzw. Begründung autonomen Handelns eingehen. In Kapitel 4.2. werden schließlich jene mentalen Phänomene aufgeführt, von denen ausgehend Autonomie ebenfalls ausschließlich dem Menschen zukommt. Der entscheidende Zusammenhang von Sprache und Handlungstheorie ist jedoch zu weitläufig, um im Rahmen dieser Arbeit näher beleuchtet zu werden; Berührungspunkte können dabei nur im Unterkapitel über die Qualia kurz angedeutet werden.

4.1. Sinnhaftigkeit von autonomen Handlungen

Zunächst muss zwischen dem Handeln, das wir meinen, wenn von einem Menschen die Rede ist, sowie dem vermeintlichen Handeln von Maschinen unterschieden werden. Im Falle von Maschinen überhaupt von Handlungen zu sprechen hieße ihnen mehr

²⁰² Vgl. Hans Lenk: Das flexible Vielfachwesen: Einführung in die moderne philosophische Anthropologie zwischen Bio,- Techno- und Kulturwissenschaften, Weilerswist 2010. Lenk führt eine mehr als 30-seitige Liste aller bekannten Charakteristika des Menschen auf. Vgl. auch Peter Janich: Was ist der Mensch?, in: Detlev Ganten et al. (Hgg.): Was ist der Mensch? Berlin 2008, 124ff.

²⁰³ Lenk 2010

zuzuschreiben als sie eigentlich leisten können, denn Handlungen bzw. Problemlösungen, die vermeintlich von Computerprogrammen entschieden werden, sind

„[...] Effekte vorgängiger, von Entwicklerseite vorgenommener Auswahlprozeduren und konstruktiver Schritte, [...] in denen zunächst einmal Probleme bestimmt werden müssen, die für eine Computerbearbeitung interessant sein könnten.“²⁰⁴

Handlungen eines Computers oder einer Maschine unterstehen der Autorität des Entwicklers somit von vorn herein. Dem menschlichen Sinn der Handlung steht das maschinelle Prinzip von Input und Output gegenüber. Die Ebenen, auf welcher die Handlungen bzw. Aktionen ausgeführt werden, ließen sich folgendermaßen einteilen:²⁰⁵

Person		
Strukturiert als	Resultat	Bezug
Handelnder	Handlung	Umgebung (Relevanzordnung)
Akteur	Operation	Umwelt (Funktionsordnung)
Agent	Funktion	Parameter-Hypervolume (Kalkülordnung)

(Abb.1 nach Gutmann 2010, 144)

Da autonome technische Systeme sich dadurch auszeichnen, dass sie verändernd in die Umwelt eingreifen können, erfüllen sie Aufgaben, die der Mensch als Akteur bewältigt. Diese Fähigkeit obliegt ihnen jedoch nur partiell, da das Handeln der Maschine als Akteur auf eine bestimmte Aufgabe reduziert ist und nicht die Gesamtheit der vom Menschen durchführbaren Operationen meint. In die darüber liegende Ebene äußert sich nun die Autonomie des Menschen in seinen Handlungen, welche beispielsweise moralischen oder pragmatischen Ursprungs sind.

Dies soll in den folgenden Unterkapiteln an der Frage nach der Sinnhaftigkeit von Handlungen näher beleuchtet werden. Autonomie spielt dabei eine zentrale Rolle: Sie wird auf der einen Seite als Ursprung und auf der anderen Seite als Rahmen und Ziel der Handlungsmotivation betrachtet. Ausgangspunkt dieser zwei Argumentationszweige ist die Grundannahme, dass Autonomie an die Definition des Menschen als Person geknüpft ist.

4.1.1. Personalität und Moralität als Leitmotiv für autonomes Handeln

Kant definiert den Menschen als sinnliches Naturwesen und vernünftiges, mit Freiheit begabtes Wesen zugleich. Autonomie, laut Kant Synonym für die Selbstgesetzgebung

²⁰⁴ Thomas Christaller / Josef Wehner: Autonomie der Maschinen – Einführung in die Diskussion, in: Thomas Christaller (Hg.): Autonome Maschinen, Wiesbaden 2003, 13

²⁰⁵ Vgl. Mathias Gutmann: Autonome Systeme und der Mensch: Zum Problem der medialen Selbstkonstitution, in: Stefan Selke / Ulrich Dittler (Hgg.): Postmediale Wirklichkeiten aus interdisziplinärer Perspektive. Weitere Beiträge zur Zukunft der Medien, Hannover 2010, 127-148

des freien Willens, manifestiert sich in dem Bereich der Definition des Menschen,²⁰⁶ der nicht der Kausalität der Natur, sondern der Freiheit unterliegt. Es ist zwar definierbar, was als Autonomie gilt (s. Kap. 2), d.h. wie sie sich auswirkt, jedoch kann die Frage wie genau sie beim Menschen bewirkt wird, nicht naturalistisch erklärt werden, da ihre Kausalität der Idee der Freiheit entspringt. Kant verknüpft Freiheit in seiner praktischen Philosophie mit dem Begriff der Person, welche „sowohl Triebfeder als auch Gegenstand der Freiheit“²⁰⁷ ist, denn Personen sind die einzigen Wesen, durch welche das moralische Gesetz gewährleistet werden kann.

Das Bewusstsein darüber, dass moralische Gesetze existieren, setzt Kant wiederum mit dem „Faktum der Vernunft“²⁰⁸ gleich. Dieses ist wiederum die Fähigkeit von Personen, „[...] sich zu ihren Neigungen und empirischen Bestimmungen verhalten zu können [...]“, sie ist ihnen „[...] gleichsam einverleibt und aufgrund dieser Einbettung eine Quelle von Motivation und Handlung.“²⁰⁹ Eine Person, die dem von ihrer Vernunft gegebenen Gesetz folgt, hat Persönlichkeit. Kant zählt Personen daher nicht nur zur Sinnen- und Naturwelt, sondern bezeichnet sie als intelligible Wesen.²¹⁰

Weil aber die Verstandeswelt den Grund der Sinnenwelt, mithin auch der Gesetze derselben enthält, also in Ansehung meines Willens (der ganz zur Verstandeswelt gehört) unmittelbar gesetzgebend ist, und also auch als solche gedacht werden muß, so werde ich mich als Intelligenz, obgleich andererseits wie ein zur Sinnenwelt gehöriges Wesen, dennoch dem Gesetze der ersteren, d.i. der Vernunft, die in der Idee der Freiheit das Gesetz derselben enthält, und also der Autonomie des Willens unterworfen erkennen, folglich die Gesetze der Verstandeswelt für mich als Imperative und die diesem Prinzip gemäßen Handlungen als Pflichten ansehen müssen.²¹¹

Damit ist die Freiheit des Menschen nicht als Indeterminismus zu verstehen, sondern tritt „[...] in der Form von Autonomie über eine eigene Gesetzmäßigkeit [...]“²¹² zum Vorschein. Der sich seiner Vernunft gewahr werdende Mensch erkennt seine Zugehörigkeit zur Sinnen- sowie zur Verstandeswelt, er leitet darüber hinaus als intelligibles Wesen, indem er von seiner Vernunft Gebrauch macht, die Gesetze seines Handelns letztlich aus der Autonomie des Willens ab.²¹³

²⁰⁶ 3. Antinomie, KpV

²⁰⁷ Georg Scherer: Person III 1-9, in: HWPh Bd. 7(P-Q) 1989, 307

²⁰⁸ Vgl. KpV, 1. Teil, 1. Buch, 56,4ff

²⁰⁹ Dieter Sturma: Was ist der Mensch? Kants vierte Frage und der Übergang von der philosophischen Anthropologie zur Philosophie der Person, In: Dietmar H. Heidemann / Kristina Engelhard (Hgg.): Warum Kant heute? Berlin 2004, 276

²¹⁰ Vgl. KpV [155]

²¹¹ GMS [453, 7 – 454, 17]

²¹² Sturma 2004, 276

²¹³ Vgl. auch GMS [427], 31-35. „Der Wille wird als ein Vermögen gedacht, *der Vorstellung gewisser Gesetze gemäß* sich selbst zum Handeln zu bestimmen. Und ein solches Verhalten kann nur in vernünftigen Wesen anzutreffen sein.“ Die Gesetzgebung des Willens wiederum ist Autonomie.

Für Kant unterliegen Sachen bzw. vernunftlose Objekte naturwissenschaftlichen Gesetzen, während die Person dazu in der Lage ist, nach vorgestellten moralischen Gesetzen zu handeln. Die von der Person mittels moralischer Gesetze erwirkte Autonomie setzt Kant wiederum mit Selbstzwecklichkeit gleich. Kant definiert „die Person“ in Abgrenzung zu „Sachen“ wie folgt:

Die Wesen, deren Dasein zwar nicht auf unserem Willen, sondern der Natur beruht, haben dennoch, wenn sie vernunftlose Wesen sind, nur einen relativen Wert, als Mittel, und heißen daher Sachen, dagegen vernünftige Wesen Personen genannt werden, weil ihre Natur sie schon als Zwecke an sich selbst, d.i. als etwas, das nicht bloß als Mittel gebraucht werden darf, auszeichnet, mithin sofern alle Willkür einschränkt (und ein Gegenstand der Achtung ist).²¹⁴

Kant folgend ist das Handeln des Menschen von Zwecken und Gründen geleitet, während „Sachen“ lediglich als Mittel zu betrachten sind. Dieter Sturma greift diese Handlungsmotivation auf, um anhand des „Raumes der Gründe“²¹⁵ zwischen der Autonomiefähigkeit von Personen und Maschinen differenzieren zu können. Er folgt der Kantischen Annahme, dass jene Wesen über Autonomie verfügen,

[...] die für Gründe empfänglich sind sowie aus Gründen verallgemeinern, differenzieren und handeln können. Mit dem Raum moralischer Gründe eröffnet sich eine eigenständige gesetzesartige Dimension, die vom Raum der Ursachen verschieden ist.²¹⁶

Ursachen gelten für die Sinnenwelt, für den Raum naturkausal erklärbarer Gesetze, während Gründe in der Vernunftwelt gedacht werden. Sturma zufolge rekurriert der Autonomiebegriff zwar auf das Handeln nach Gesetzen, der Unterschied liegt jedoch bei dem Ausführenden der Handlung: eine Person handelt erst dann autonom, wenn sie „[...] ein vorgestelltes Gesetz zur Anwendung bringt [...]“²¹⁷. Die Gesetze sind nicht in einer Person vorhanden, sondern werden von dieser abstrahiert bzw. begründet. Selbstbewusste Wesen gelten als Personen, sofern sie folgende Bedingungen von Autonomie erfüllen:

[...] für Gründe empfänglich und im Stande sein, von ihren Neigungen abstrahieren sowie aus Gründen verallgemeinern, differenzieren und handeln zu können.²¹⁸

Auf diesem Wege erfolgt die Abgrenzung der Autonomie des Menschen über eine naturkausale Nicht-Erklärbarkeit. Sturma weist beispielsweise auch darauf hin, dass die auf Personalität beruhende Definition von Autonomie lediglich deshalb auf den

²¹⁴ GMS [428] 1-8

²¹⁵ Vgl. Sturma 2004

²¹⁶ Sturma 2004, 277

²¹⁷ Dieter Sturma: Über Personen, Künstliche Intelligenz und Robotik, in: Thomas Christaller (Hg.): Autonome Maschinen, Wiesbaden 2003, 40

²¹⁸ Ebd.

Menschen anwendbar ist, da „[...] uns bislang Personen nur in der Gestalt von Menschen bekannt“²¹⁹ sind. Ferner schreibt er, „[...] Personen müssen nicht Individuen der Spezies *homo sapiens sapiens* sein [...]“, und doch ist „eine spezifische Form von Körperlichkeit eine unabdingbare Voraussetzung für personales Leben.“²²⁰ Das Spezifische an Personen besteht darin, dass sie sich aufgrund ihres Selbstbewusstseins in verschiedener Weise zu sich Selbst und zu anderen Körpern in ihrer Umwelt verhalten, denn „[d]er menschliche Eigenkörper (*corps propre*) bzw. der Leib repräsentiert bereits Haltungen von Personen zu ihrer Erfahrungswelt.“²²¹ Diese Haltung hebt die Sonderstellung des Menschen als Sinnes- und Vernunftwesen hervor, das sich mittels der Selbstzuschreibung von Personalität einen moralischen Rahmen für seine Handlungsmotivation geschaffen hat. Dabei dient Autonomie dazu, wie Sturma feststellt, „[...] Gesetze für Maximen – worunter Lebensregeln zu verstehen sind – festzulegen.“²²² Erfüllten künstliche Systeme diese Voraussetzungen von Personalität, so könnten sie Sturma zufolge mit Recht als Personen bezeichnet werden.

Diese von Kant abgeleitete Anbindung von Autonomie an den Personenbegriff ist zwar schlüssig, es bleibt jedoch fraglich inwieweit es überhaupt erforderlich ist den Personenbegriff an eine bestimmte Form von Körperlichkeit zu knüpfen, um sie zum Gegensatz von Mensch und Maschine zu erheben. Die Definition von Autonomie, die Sturma in dem bisher zitierten Essay bietet, kommt ebenso gut ohne den Personenbegriff aus und ist allein als Gegensatz von vernunftfähigen Menschen zu vernunftlosen Maschinen lesbar:

Der Begriff der Autonomie lässt sich nur auf eine Lebensform anwenden, die sich dem Raum der Gründe öffnet. [...] Menschen [sind] im Normalfall Personen [...], die sich selbstständig im semantischen Raum der Gründe und Handlungen bewegen.²²³

Sturma geht von einem Menschenbild aus, in dem der Mensch aufgrund bestimmter Fähigkeiten äquivalent mit ‚der Person‘ ist, welche die Maschine aufgrund des Fehlens eines großen Teils der Fähigkeiten nicht ist, bei Erfüllung aber sein könnte; da aber das Konstrukt ‚Person‘ überhaupt erst das Resultat von Selbsterfahrung und Erfahrung der Lebenswelt eines Menschen sein soll,²²⁴ müsste der Maschine Personalität jedoch a

²¹⁹ Sturma 2003, 40

²²⁰ Ebd.

²²¹ Ebd.

²²² Ebd., 39

²²³ Sturma 2003, 49. Vgl. auch folgendes Zitat in ebd. 43: „Der Raum der Gründe ist Ausdruck des folgenreichen Sachverhalts, dass Menschen über einen langen Zeitraum ihre Naturgeschichte in Kulturgeschichte transformiert haben. In ihm sedimentiert sich formal und inhaltlich die Entwicklung der Begriffe, die den epistemischen, moralischen und ästhetischen Eigensinn menschlicher Fähigkeiten und Eigenschaften konstituiert sowie einen neuen Rahmen für Denk- und Verhaltensweisen erzeugt.“

²²⁴ Vgl. Sturma 2003, 40: „Unter normalen biologischen und sozialen Verhältnissen entwickeln sich menschliche Individuen zu Personen[...]“.

priori verwehrt bleiben. Nach Sturmas eigenen Prämissen kann Personalität somit nie zu einem Kriterium werden, an dem man die Äquivalenz von Mensch und Maschine festmachen könnte, da er Maschinen von vornherein die Zugehörigkeit zum „Raum der Gründe“ abspricht.

Zuzustimmen ist dennoch Sturmas Unterscheidung von „schwachen“ und „dichten“ Beschreibungen menschlicher Lebensformen:²²⁵

Während eine schwache Beschreibung mit der Zusammenstellung von empirischen Daten auskommt, die aus der Perspektive des äußeren Beobachters gewonnen werden können, bezieht die dichte Beschreibung Bestimmungen mit ein, die konstitutiv für die Teilhabe an der Lebensform sind und infolgedessen einem externen Standpunkt unzugänglich bleiben.²²⁶

Autonomie ist insofern als ein spezifisch menschliches Charakteristikum anzusehen, dessen Bestimmung bei künstlichen nachgebildeten Systemen und Computerprogrammen keine Anwendung finden kann, da die Grundvoraussetzungen bereits solche sind, die keiner wissenschaftlich begründbaren Ursache folgen, sondern mit epistemischen oder moralischen Gründen erklärt werden. Der Handlungsraum der Maschine hingegen wird stets vom Menschen präformiert. Da sie Vorgänge syntaktischer Art umsetzen sind sie im „logischen Raum der Gründe“²²⁷ aktiv. Damit beherrschen sie jedoch nur einen partiellen Teil der menschlichen Lebensform, denn allein mit „automatisierten Zügen“ ist kein Verständnis von Semantik oder die Fähigkeit zu Selbstreferenzialität gewährleistet.²²⁸ Das heißt, jene epistemisch kontextualisierten Begriffe, derer sich eine Person im „Raum der Gründe“ gewahr wird, werden im Zusammenhang der Robotik usw. lediglich als „schwache Beschreibung“ genutzt, es handelt sich lediglich um „als-ob-Autonomie“; „als-ob-Intelligenz“ oder auch „als-ob-Intentionalität“. Diese klare Unterscheidung ist gerade deshalb von Bedeutung, da im Umfeld von KI und Robotik die Anforderungen fälschlicherweise an ein verschlanktes, von Input und Output geprägtes Menschenbild orientiert werden. Partielle Fähigkeiten werden auf das Gesamte übertragen.

4.1.2. Zur Pragmatik der Autonomie des Handelns

Der Mensch handelt aus pragmatischen Gründen autonom. Autonomie bedeutet in diesem Sinne das Erwägen bzw. Ausschöpfen gegebener Handlungsmöglichkeiten und bildet zugleich den Rahmen und das Ziel der Handlungsmotivation. Der autonome Mensch ist in der Lage, sich eine Vielfalt an Handlungsräumen zu schaffen und in diesen zu interagieren. Diese Position entspringt im Grunde der in Kapitel 2.3 erläuterten aristotelischen, teleologisch orientierten Handlungsmotivation.

²²⁵ Siehe Sturma 2003, 41

²²⁶ Ebd.

²²⁷ Ebd., 43

²²⁸ Ebd.

Der technologische Fortschritt spielt dabei eine wesentliche Rolle, da der Mensch Technologien stets dahingehend weiterentwickelt, um diese zur Problemlösung, zur Unterhaltung oder als spezialisierte Werkzeuge zu nutzen.²²⁹ Während sich der Mensch hierdurch die Kommunikation und Interaktion mit anderen Menschen erleichtert, haben die jeweiligen Technologien selbst keine „[...] Teilhabe an sinnförmig vermittelten Interaktions- und Kommunikationsbeziehungen [...]“²³⁰. Ihre sogenannte Vergesellschaftung besteht darin, dass sie vom Computerspezialisten mittels Programmiersprache dem Menschen nähergebracht werden. Dies wiederum geschieht einerseits durch anthropomorphe Erklärungen der Technik, und andererseits muss die Technik selbst, um sich durchsetzen zu können derart gestaltet sein, dass sie den Vorstellungen und Erwartungen des Menschen entspricht. Computertechnik ist somit an die Entwicklung der Gesellschaft gebunden und unterliegt stets ihren Entscheidungen:

Veränderungen der Gesellschaft, die oft mit technischen Entwicklungen erklärt werden, sind also immer auch Ausdruck von Entscheidungen und Zielsetzungen, die in der Computertechnik Gestalt annehmen und schleifenförmig in die Gesellschaft zurücklaufen.²³¹

Computertechnik verhält sich somit nicht anders als andere zuvor entwickelte Techniken, sie ist vielmehr Fortführung, Spezialisierung und Steigerung. In diesem Sinne vollzieht sich ein allmählicher Übergang von der Automatisierung zur Autonomisierung, wobei der Mensch weiterhin dieselben Ziele verfolgt. Es geht darum,

[...] dass Technik im Wesentlichen der Steigerung bereits vorgängiger menschlichen Vermögen (Denken, Kommunikation, Mobilität) dient. In ihr zeichnet sich der Mensch durch Bewusstsein, Willensfreiheit und Nichtvorherbestimmbarkeit aus, während die Technik funktionieren muss, das heißt immer wieder exakt die ihr bestimmten Vorgänge zu reproduzieren hat.²³²

Insofern ist der Mensch in seinem Handeln autonom, seine Handlungen sind nicht determiniert, sondern können bewusst gesteuert, d.h. Handlungen können trotz gleicher Situationen variiert, Entscheidungen verändert werden. Dieses Merkmal von Autonomie ist ein wesentliches, das Maschinen nicht zukommt.

Um mit dem Begriff der Autonomie sowohl dem menschlichen Verhalten als auch dem der Maschine gerecht zu werden, muss Christaller und Wehner zufolge entweder den Sinngehalt menschlicher Autonomie im Hinblick auf die damit verbundenen Besonderheiten menschlicher Intelligenz derart vereinfacht werden, dass die Autonomie-Definition dann selbst das für das „Verhalten“ von Maschinen gilt. Oder aber es wird mit dem Ausdruck Autonomie der gesamte Bedeutungsgehalt menschlicher Fähigkeiten

²²⁹ Vgl. Christaller / Wehner 2003, 19

²³⁰ Christaller / Wehner 2003, 14

²³¹ Ebd., 15

²³² Ebd., 16

unterstellt, den die Maschine aufgrund ihrer Freiheit von Sinn, Bewusstsein, mentalen Zuständen usw. nicht leisten kann.²³³

4.2. Mentale Phänomene, Qualia

In diesem Kapitel fasse ich sämtliche Begriffe zusammen, die aus dem Körper-Geist-Dualismus als alleinstellende Merkmale des Menschen hervortreten.

Ungeachtet der Frage nach der Richtigkeit von interaktionistischem oder nicht-interaktionistischem Dualismus, schicke ich an dieser Stelle voraus, dass die Frage an sich schon suggeriert, dass die künstliche Nachbildung eines auch biologisch bzw. biochemisch noch nicht gelösten philosophischen Problems zumindest auf den nichtinteraktionistischen Dualismus rekurrieren müsste, um das entstehende Produkt mit dem Menschen vergleichen zu können. Wie aber sollen die Eigenschaften eines nicht-menschlichen Systems mit denen eines Menschen verglichen werden, wenn auf einer übergelagerten Ebene nicht zwischen den Dualismusformen und darüber hinaus zwischen Dualismus und Monismus entschieden werden kann?

Vor diesem Hintergrund erörtert John Searle in seiner Vorlesungsreihe *Minds, Brains and Science*²³⁴ Gründe für die Sonderstellung des Menschen, die sich trotz der durch den Menschen bewusst herbeigeführten zunehmenden Technisierung nicht verändert hat. Noch immer sind die Diskussionen um den Körper-Geist-Dualismus nicht überwunden, denn gewisse geistige Vorgänge scheinen einfach nicht in das naturwissenschaftliche Weltbild zu passen. Die vier dem Menschen eigentümlichen mentalen Phänomene gelten laut Searle ausschließlich für den Menschen und sind nicht auf die Maschine übertragbar: Hierzu gehören (a) Bewusstsein (*consciousness*), (b) Intentionalität (*intentionality*), (c) Subjektivität mentaler Zustände (*subjectivity of mental states*) sowie (d) mentale Verursachung (*mental causation*).²³⁵ Searle definiert sie folgendermaßen:

[...] namely, mental states are biological phenomena. Consciousness, intentionality, subjectivity and mental causation are all a part of our biological life history, along with growth, reproduction, the secret of bile, and digestion.²³⁶

Mentale Zustände sind somit als biologische Phänomene Teil des biologischen Lebens, was zunächst ihre naturkausale Erklärbarkeit suggeriert, jedoch ihre Zugehörigkeit zur naturkausal-nicht-erklärbaren Welt nicht ausschließt. Das bedeutendste, weil meist diskutierte und noch ungelöste Problem ist das des Bewusstseins. Searle kommt zu dem

²³³ Christaller / Wehner 2003, 16

²³⁴ John R. Searle: *Minds, brains and science* (The 1984 Reith Lectures), Cambridge (Massachusetts) 2003

²³⁵ Vgl. Searle 1984, 16

²³⁶ Searle 1984, 41. Ähnliches formuliert auch Sturma: „Wesentliche Elemente von Autonomie im nicht-übertragenen Sinne sind Bewusstsein, Intelligenz, Intentionalität, Wille, Reflexion, Entscheidung, Handlung und Präsenz im Raum der Gründe.“ (Sturma 2003, 49)

Schluss menschliches Leben sei ohne Bewusstsein nicht möglich, da hierdurch andere spezifisch menschliche Eigenheiten erst hervorgebracht werden.

Consciousness is the central fact of specifically human existence because without it all of the other specifically human aspects of our existence – language, love, humour, and so on – would be impossible.²³⁷

Folglich wäre eine rein physische Welt, in der es kein Bewusstsein gibt, bedeutungslos, ihre Existenz wäre nicht feststellbar. Searle wirft zwar die Frage auf, wie Bewusstsein im Gehirn überhaupt entsteht, doch gibt er keine Antwort hierauf. Er geht zur Erläuterung des zweiten Phänomens über, zur Intentionalität. Sie ist die teleologische Funktion, durch welche die mentalen Zustände des Menschen auf etwas bezogen werden: bspw. Absichten, Hoffnungen oder Wünsche. Intentionalität bezieht sich sowohl auf bewusst als auch unbewusst Erlebtes. Ebenso schwer fügt sich Subjektivität in das wissenschaftliche Weltbild der neutralen Objektivität. Searle verweist hierdurch auf die Unübertragbarkeit personaler mentaler Zustände auf einen anderen Menschen. Mentale Zustände werden stets subjektiv wahrgenommen, seien es Schmerzen oder die Sicht auf ein Weltbild. Zum Problem der mentalen Kausalität, die auf die „Körper-Geist-Problematik“ anspielt, fragt Searle: „But if our thoughts and feelings are truly mental, how can they affect something physical?“²³⁸ Sein Lösungsvorschlag ist eine Verschmelzung des naiven Physikalismus mit dem naiven Mentalismus. Aus diesen Ausführungen resultiert folgende These:

Mental phenomena, all mental phenomena whether conscious or unconscious, visual or auditory, pains, tickles, itches, thoughts, indeed, all of our mental life, are caused by processes going on in the brain.²³⁹

Sie alle werden in einer Mikrostruktur verursacht und wahrgenommen, und dieses Verhältnis zueinander ist dasjenige von „mind to brain“, von Geist zu Gehirn.²⁴⁰ Das Gehirn und sämtliche mentale Zustände sind Searle zufolge biologisch erzeugt. An Oberfläche, einer sogenannten Makroebene, wird schließlich dasjenige sichtbar bzw. wahrnehmbar, was im Gehirn auf Mikroebene durch bestimmte biochemische Prozesse verursacht wird.²⁴¹

Um nun zu klären, ob Computer denken können, versucht Searle die vorgestellten Punkte auf das Funktionieren eines Computerprogramms zu übertragen und kommt zu

²³⁷ Searle 1984, 16

²³⁸ Searle 1984, 17

²³⁹ Ebd., 18

²⁴⁰ Vgl. ebd. 1984, 18 sowie 22: „[...] there is no difficulty in accounting for the relations of the mind to the brain in terms of the brain's functioning to cause mental states.“

²⁴¹ Vgl. ebenso Eduard Zwierlein: Künstliche Intelligenz und Philosophie. Zur Debatte um J. R. Searle's Einwände gegen harte KI-Versionen, in: *Journal for General Philosophy of Science*, 21 (1990), 351

dem Ergebnis, dass dies unmöglich ist.²⁴² In *Minds, Brains and Science* führt Searle die Erläuterung der vier Phänomene weiter und entwickelt vier Prämissen:

1. Brains cause minds.[...]
2. Syntax is not sufficient for semantics. [...]
3. Computer programs are entirely defined by their formal, or syntactical, structure. [...]
4. Minds have mental contents; specifically, they have semantic contents.²⁴³

Diese führen zu der Konklusion, dass ein Computerprogramm nie ausreichen wird, „to give a system a mind. Programs [...] are not by themselves sufficient for having minds.“²⁴⁴ Doch was bedeutet dies für den Menschen selbst? Was lässt sich nun aus Searles Standpunkt für die Frage nach der Exklusivität des Autonomiebegriffs für den Menschen ableiten? Um menschliche und künstliche Denkfähigkeit gegenüberzustellen, hebt Searle meist die zweite Prämisse hervor. Während das Computerprogramm lediglich auf einer syntaktischen Ebene arbeitet, ist der Mensch in der Lage, denselben syntaktischen Inhalt semantisch aufzufassen.

Solange es keine wissenschaftliche biologische Klärung der mentalen Zustände gibt, sondern wir eine biochemische Reaktion im Gehirn vermuten und nur modellhaft von Intentionalität oder der Bewusstseinsproblematik sprechen, können diese auch nicht ausgeblendet werden. Daher dürfen, sofern die vier beschriebenen mentalen Phänomene nicht ausschließlich wissenschaftlich erklärbar sind, die dem Menschen eigentümlichen Zuschreibungen nicht preisgegeben werden. Searles Argument von der biologischen Erklärbarkeit scheint einen währenden Vorteil zu haben. Es bleibt bis heute anwendbar, selbst nachdem sich die KI-Forschung von der herkömmlichen von-Neumann-Programmierung²⁴⁵ losgelöst und auf die Entwicklung von Synapsen-Prozessoren kapriziert hat mit dem Ziel „[...] to enable true large mammalian brain-scale networks [...]“²⁴⁶. Wie soll man herausfinden, ob ein Synapsen-Prozessor vermeintlich einen mentalen Zustand hervorbringen kann, wenn man keine biologische Erklärung für die

²⁴² Searle 1984. 39

²⁴³ Searle 1984, 39

²⁴⁴ Ebd. An dieser Stelle ließe sich „mind“ gleich mit unterschiedlichen mentalen Zuständen übersetzen wie etwa Sinn, Verstand, Geist, Seele.

²⁴⁵ Einen Überblick über die verschiedenen von-Neumann-Architekturen sowie deren Schwächen und Weiterentwicklungsmöglichkeiten liefert John Backus 1978 in einem Paper, siehe: John Backus: Can Programming Be Liberated from the von Neumann Style? A Functional Style and Its Algebra of Programs, in: *Communications of the ACM* 21/8 (1978), 613-641.

²⁴⁶ Jae-sun Seo et al.: A 45nm cmos neuromorphic chip with a scalable architecture for learning in networks of spiking neurons, in: *Custom Integrated Circuits Conference (CICC), 2011 IEEE*, Sept. 2011, 1. In der Entwicklung von Synapsenprozessoren, die über die gewöhnliche Funktion des Speicherns und Verarbeitens von Daten hinausgehen, verzeichnen Forscher des Programms DARPA seit 2009 Erfolge. Von der Funktionsweise des menschlichen Gehirns inspiriert, funktionieren Teile dieser Prozessoren wie ‚digitale Neuronen‘ und können dadurch auch Informationen verarbeiten, für die sie nicht programmiert wurden. (Siehe Seo et al. (2011) sowie DARPA SyNAPSE Program – Neuromorphic Adaptive Plastic Scalable Electronics, <http://www.artificialbrains.com/darpa-synapse-program>)

menschliche Entsprechung hat, mit der man diesen vergleichen könnte? Man liefe weiterhin Gefahr, von einem *Chinese Room*²⁴⁷ getäuscht zu werden. In diesem Sinne müssen wir, wenn von Maschinen die Rede ist, auch weiterhin von einer Simulation menschlicher Eigenschaften ausgehen und damit auch von einer „als-ob-Autonomie“.

Es sollte jedoch angemerkt werden, dass Searles Thesen, obwohl sie vollkommen für die Exklusivität der Autonomie des Menschen sprechen, im Grunde ein Spiel auf Zeit mit der biochemischen Forschung sind. Searle fasst etwa den Zusammenhang von Denk- und Handlungsprozessen wie folgt zusammen: „The whole apparatus works because it is realized in the brain[...]“²⁴⁸ Insofern wäre auch die Autonomie des Menschen das Ergebnis eines Vorgangs im Gehirn. Indem er sämtliche mentale Zustände und Vorgänge auf das Gehirn reduziert, erklärt er sie zu dessen Produkten; jedoch kann ihre naturwissenschaftliche Entschlüsselung dadurch nicht mehr ausgeschlossen werden.

Zudem widerspricht sich Searle in einer späteren Veröffentlichung selbst, als er Gehirne als Maschinen bezeichnet (was sie berechenbar macht) und damit die oben aufgeführten vier Prämissen schwächt, die dem Computer Denkfähigkeit absprechen sollen. Prämisse 3 (*Computer programs are entirely defined by their formal, or syntactical, structure.*)²⁴⁹ ist das Beispiel für die Funktionsart einer Maschine. Fügt man hier noch die Prämisse hinzu, Gehirne seien im Grunde Maschinen, so trägt die Konklusion „Programs [...] are not by themselves sufficient for having minds“²⁵⁰ eine ganz neue Aussage: Da aus der neuen Prämissenzusammensetzung folgt, dass Gehirne und Maschinen dieselben Funktionen hervorbringen, wären sie demzufolge beide „not by themselves sufficient for having minds“²⁵¹. Damit wird Searles Theorie für Vertreter der starken KI, wie etwa Ray Kurzweil, angreifbar.²⁵² Während Kurzweil Searle mangelndes Verständnis für das technisch Realisierbare attestiert und seine Theorie für tautologisch und widersprüchlich hält, wirft Searle seinem Kontrahenten vor, er würde ausgehend von bisherigen Erfolgen in der Computertechnologie auf die Existenz-

²⁴⁷ Mit dem Gedankenexperiment *Chinese Room* macht Searle darauf aufmerksam, dass Computerprogramme lediglich syntaktische und keine semantischen Inhalte verarbeiten oder verstehen können. Searle imaginiert einen Raum, in dem sich eine Person aufhält, die lediglich Englisch, jedoch kein Chinesisch versteht. Ohne sich zu erkennen zu geben, erhält diese Person ein Papier mit Aufgaben in chinesischen Schriftzeichen, welche sie mithilfe der im Raum befindlichen Wörterbücher löst und auf der anderen Seite des Raumes an eine dritte Person weiterreicht. Searles Einwand ist, dass die dritte Person allein aufgrund der zu Papier gebrachten Übersetzung, also des Outputs, zu der Annahme verleitet wird, die in dem Raum befindliche Person sei der chinesischen Sprache mächtig. Siehe John Searle: *Minds, brains, and programs*, in: *The Behavioral and Brain Sciences*, 3 (1980a), 417-457.

²⁴⁸ Searle 1984, 64

²⁴⁹ S.o.

²⁵⁰ Searle 1984, 39

²⁵¹ Vgl. ebd.

²⁵² Vgl. hierzu das Band Jay W. Richards (Hg.): *Are we spiritual machines? Ray Kurzweil vs. the critics of strong AI*, Seattle/Washington 2002

möglichkeit von starker KI schließen und damit seine Szenarien für realisierbar halten.²⁵³ Interessant ist an dieser Debatte, dass sowohl Kurzweil als auch Searle die Möglichkeit bzw. Unmöglichkeit von KI aus ihrem gesicherten Wissen über Teilbereiche der KI zu extrapolieren versuchen: Searle über das *Chinese Room*, Kurzweil über seine eigenen Erfolge, beispielsweise in der optischen Spracherkennung. Dies ist jedoch ein wichtiger Punkt, denn in der Tat beweist das *Chinese Room* höchstens, dass starke KI im Sinne von Spracherkennung nicht möglich ist.

4.3. Bewertung und Zusammenführung

Die in Kapitel IV aufgeführten Thesen stützen die Position, dass Autonomie ein Kriterium ist, das lediglich dem Menschen zukommt. Sofern man Autonomie als ein der Moralphilosophie entspringendes Charakteristikum des Menschen betrachtet, ist diese als exklusiv nur für den Menschen gültig zu betrachten. Dies wurde kontinuierlich durch die Abgrenzung der Autonomie des Menschen von technischen Systemen vorgenommen. Die moralphilosophische Autonomiedefinition an sich unterscheidet sich von der Definition technischer Autonomie derart, dass sie durch diese weder bereichert noch angegriffen wird.

Das letzte Kapitel zeigt, dass es hinsichtlich der Qualia tatsächlich noch derart viel Erklärungsbedarf zu ihrem Wirken, ihrem Ursprung und dem Zusammenhang zwischen Körper und Geist gibt, dass von einer Äquivalenz von Mensch und Maschine in dieser Hinsicht nicht ausgegangen werden kann. Bereits in Kapitel 3 ging aus der Differenzierung der KI-Ansätze nach Russel und Norvig hervor, dass die eigentlichen Ziele nicht auf eine identische Nachbildung des Menschen abzielen. Dennoch sind autonome Systeme und Roboter zunehmend Bestandteil der unmittelbaren Umwelt des Menschen; d.h. der Handlungsraum des Menschen, in welchem er aus moralischen und epistemischen oder pragmatischen Gründen handelt, ist sogleich der Aktionsraum der Maschine.

²⁵³ Vgl. insbes. Richards 2002, 76 sowie 170

Manchmal dringt eine sehr komplizierte Idee in derart vereinfachter Form ins öffentliche Bewußtsein, daß sie nunmehr eine Karikatur des Originals darstellt; aber dieser bloße Schatten ist sehr wohl in der Lage, den Begriff entscheidend zu verändern, den die Allgemeinheit von der Realität hat.

(Weizenbaum 1977, 207)

5. Erweiterung der Autonomie des Menschen

In diesem Kapitel wird das Wechselverhältnis zwischen Autonomie des Menschen und technischer Autonomie betrachtet sowie ihre Interaktion und Einbettung im Gesamtkontext von Autonomie und autonomem Handeln. Es wird dabei untersucht, inwieweit technische Autonomie dabei lediglich Sprachspiel ist und inwieweit sie den Handlungsraum des Menschen derart beeinflusst, dass hierdurch von einem neuen, mensch-maschinen-hybriden Menschenbild ausgegangen werden kann. Dabei soll aufgezeigt werden, dass durch das grundsätzliche Zuschreiben von Fähigkeiten des Menschen an sich selbst sowie an die Maschine die Möglichkeit offen steht und genutzt werden sollte, autonome Maschinen als Erweiterung der Autonomie des Menschen zu begreifen – was jedoch nur unter der Festlegung für mensch- und maschinenspezifische Autonomiegrade gelingen kann.

Hierzu zeige ich zunächst sprachliche und inhaltliche Abgrenzungsprobleme auf, die zwangsläufig auftreten, wenn über Technologien gesprochen wird, denn ein wesentlicher Teil dessen, was wir von der Funktionsweise autonomer Systeme wissen, hängt von dem ab, was wir ihnen sprachlich zuschreiben.

Im zweiten Teil dieses Kapitels wird technische Autonomie als Erweiterung der Autonomie des Menschen im Sinne der Schaffung neuer Handlungsmöglichkeiten aufgefasst. Dies ist zum einen an der historischen Entwicklung nachvollziehbar und zum anderen durch die bewusste Etablierung zukunftsfähiger Menschenbilder, in denen der Mensch autonome (Teil-) Technologien bewusst seiner eigenen Handlungsgewalt unterstellt und sich Autonomie weiterhin als Alleinstellungsmerkmal wahr.

Die Autonomiebegriffe, die in diesem Kapitel benutzt werden, unterscheiden sich wie folgt:²⁵⁴

- (1) Zwecksetzungsautonomie: Die Autonomie des Menschen, Technologien einen Zweck, zu dem sie gebaut werden, vorzugeben.
- (2) Quantitative Autonomie: Autonomie als Vielfalt der Handlungsmöglichkeiten.
- (3) Qualitative Autonomie: Autonomie in moralischer Hinsicht; begriffliche Abgrenzung zur quantitativen Autonomie.

5.1. Aufgabenteilung und Zuschreibung von Autonomie

Die Realisierung neuer Technologien sowie bestimmter Fähigkeiten, wie eben der Autonomie, vollzieht sich sprachlich meist viel schneller als dass die Technologien tatsächlich technisch umgesetzt werden können.

Im ersten Unterkapitel soll festgestellt werden, inwiefern unser alltagssprachlicher Gebrauch von Metaphern die Vorstellung dessen, was unter autonomen Technologien verstanden wird, beeinflusst. Im zweiten Unterkapitel wird die paradoxe Situation angesprochen, in der sich der Mensch geschlossenen autonomen Systemen gegenüber

²⁵⁴ (1) ist angelehnt an Gutmann et al. 2011, während (2) und (3) von der Verfasserin zur begrifflichen Unterscheidung eingeführt werden.

befindet: Es scheint, als seien diese „black boxes“, in deren Funktionsweise der Mensch keinen Einblick mehr hat. Ich werde die These befürworten, dass über solche Systeme, trotz ihrer inneren Geschlossenheit, von außen bestimmt werden kann, indem Kriterien für die Reichweite der technischen Autonomie vorgegeben werden.

5.1.1. Missverständliche Metaphorik?

Joseph Weizenbaum beschreibt in seinem Werk *Die Macht der Computer und die Ohnmacht der Vernunft* Computer sowie sämtliche andere Modelle künstlicher Intelligenz als einen Bereich, zu dem sich der Mensch üblicherweise über Metaphern Zugang verschafft, ohne ein klares Wissen von dem zu Grunde liegenden technologischen Kontext zu besitzen.²⁵⁵

Der Öffentlichkeit ist weitgehend unklar – obwohl sie vom Gegensatz fest überzeugt ist -, was es heißt, daß prinzipiell jedes effektive Verfahren von einem Computer durchgeführt werden kann. Da der Mensch, die Natur und selbst die Gesellschaft mit Verfahren arbeiten, die zweifellos in einer anderen Hinsicht »effektiv« sind, folgt daraus, daß ein Computer zumindest den Menschen, die Natur und die Gesellschaft in allen verfahrensmäßigen Aspekten imitieren kann. Und damit ist alles (noch einmal dieses Wort!) zumindest potentiell in der Sprache von Computermodellen und –metaphern verständlich. In der Tat, auf der Basis dieser ungerechtfertigten Verallgemeinerung der Worte »effektiv« und »Verfahren« ist auch das Wort »Verständnis« neu definiert. Für alle, die sich ganz im Bann der Computermetapher befinden, bedeutet X zu verstehen, daß man in der Lage ist, ein Computerprogramm zu schreiben, das X realisiert.²⁵⁶

Die theoretischen Grundlagen von KI-Technologien sind Weizenbaum zufolge derart kompliziert, dass sie für wenige Menschen verständlich sind. Die Erläuterung ihrer Funktionsweise mittels Metaphern, wie etwa „Computer“, (genauso verhält es sich mit den metaphorischen Gebrauch von Begriffen wie „denken“, „effektives Verfahren“, „Intentionalität“) ist zunächst sinnvoll, da mittels dieser aufgrund von Ähnlichkeiten ein Mindestmaß an Verständnis zustande kommt. Ausgeblendet wird jedoch meist, dass solche Eigenschaften von Maschinen nicht auf dieselbe Art und Weise entstehen oder funktionieren wie beim Menschen. In Weizenbaums Beispiel der „effektiven Verfahren“ ist es ohne das zugehörige Fachwissen kaum abschätzbar, welcher Output an Rechenleistung ein Computerprogramm erbringen kann. Während der Mensch aufgrund der Metapher die Leistung des Artefakts an den für den Menschen geltenden Definitionen misst, kann das Artefakt selbst entweder in der Lage sein, die erwartete Leistung zu überbieten (wie im Fall der „effektiven Verfahren“) oder es werden Fähigkeiten unterstellt, die technologisch nicht umsetzbar sind.

Das auf metaphorischer Ebene erzeugte sprachliche Verständnis führt mitunter zu der Überzeugung, dass dieses zwangsläufig technisch realisierbar sei. Die beim Vorbild

²⁵⁵ Weizenbaum 1977, 210

²⁵⁶ Weizenbaum 1977, 210f

Mensch abgeschauten Fähigkeiten, welche technisch umgesetzt werden, werden jedoch zunächst auf ein Modell reduziert. Die Vergleichbarkeit reduziert sich somit auf Mensch-Modell und technologisches Mensch-Modell. In wissenschaftlichen Kontexten ist die Metapher somit auf der ersten Sprachebene „verständnisorientiert“, wohingegen auf einer zweiten Sprachebene „die Verwendung von Modellen der Normierung der Sprachpraxis, die Konstruktion einer Normsprache dient.“²⁵⁷

Ebenso verhält es sich mit dem Begriff „autonom“, welcher ebenfalls zu jenen Metaphern gehört, die eine Ähnlichkeit zu bestimmten menschlichen Leistungen suggerieren. Geht man hierbei vom Vorbild Mensch aus, so wäre die sich aus der Metapher ergebende zugehörige Modellierung in zweierlei Hinsicht zu vollbringen: Autonom -sein einerseits durch das ‚sich-autonom-Denken‘ und andererseits durch das ‚autonom-Handeln‘. Jedoch wird, wie sich in Kapitel 3 gezeigt hat, das Attribut „autonom“ einer Technologie nicht in epistemischer Weise definiert sondern lediglich mit der selbstständigen Schaffung von Handlungsmöglichkeiten gleichgesetzt. Dieser Umstand ist nach Gutmann und Rathgeber als eine modellhafte Reduktion des autonomen Handelns des Menschen zu verstehen:

Ausgegangen wird nämlich zunächst von handelnden Personen, die die zu modellierende Leistung erbringen, und im zweiten Schritt wird ein Sprachebenenwechsel vollzogen (wie wir ihn von der körperlichen Modellierung leiblicher Verhältnisse her kennen), wobei nun nicht mehr von diesen Leistungen selber, sondern von Operationen die Rede ist, welche in einer gewissen Reihenfolge und Abhängigkeit voneinander zu vollziehen wären.²⁵⁸

Dieses Zitat verdeutlicht die Ebene, auf welcher autonome Artefakte – mit kognitiven Fähigkeiten versehene Technologien generell – agieren: Während das aus der Metapher hervorgegangene Modell einer bestimmten typisch menschlichen Fähigkeit nur einen Teilbereich dessen aufzeigt, was semantisch unter dem Begriff selbst verstanden wird, ist das auf Grundlage des Modells entstandene Artefakt durchaus in der Lage, die Leistung des Menschen in einem bestimmten Teilbereich zu übertreffen. Mit „handelnden Personen“ wird hier Bezug genommen auf jenes als „Hervorbringung des Denkens“ bezeichnete Handeln; dieses ist ein Teilbereich menschlicher Handlungsräume, nämlich jener, der von Agenten ausgeführt in Funktionen und von Akteuren in Operationen mündet.²⁵⁹ Den Hintergrund hierzu erläutert Gutmann in Anlehnung an König wie folgt:

²⁵⁷ Gutmann, Mathias / Benjamin Rathgeber: Zur Leistung und Funktion von Metaphern, in: Michael Bölker et al. (Hgg.): Menschenbilder und Metaphern im Informationszeitalter, Berlin 2010, 40

²⁵⁸ Gutmann 2010, 143

²⁵⁹ Vgl. Gutmann 2010, 143f. Am Beispiel des Denkens: „Denken als Leistung eines technischen Systems wäre also nur insofern Denken, als es um die Funktionalisierung eines als Hervorbringen (von Argumenten, Rechenschritten oder Spielzügen o.Ä.) beschriebenen menschlichen Handelns zu tun ist.“

Das Verstehen dessen, was mit »Denken« stellvertretend für kognitive Ausdrücke formuliert ist, hängt an der eigentümlich generischen Funktion. Dies konstituiert nicht nur in gerader Rede dasjenige, was wir oben als gegenständliches Verhältnis im Sinne der Askription dargestellt hatten, vielmehr etabliert sich in ihr und durch sie ein Selbstverhältnis des sich als Denkenden Verstehenden, das nun nicht mehr einfach in die Konjunktion zweier Beschreibungen aufgespalten werden kann, deren eine den Denkenden und deren andere das Gedachte zum Gegenstand hat.²⁶⁰

Für die Unterscheidung von autonom sein und autonom handeln ist dieses Zitat in zweifacher Hinsicht relevant. Was hier mit „Selbstverhältnis des sich als denkenden Verstehenden“ beschrieben wird, gilt ebenfalls für die Fähigkeit des Menschen, sich selbst als autonomes Wesen zu begreifen; d.h. autonom sein konstituiert sich im sich-autonom-Denken und Handeln, wobei das autonome Handeln stets mit dem autonomen Denken zusammenfällt.²⁶¹ Aus der bloßen, ersichtlichen Handlung werden hingegen die Kriterien für Modelle des autonomen Funktionierens von technologischen Systemen abgeleitet. Somit ist autonomes Handeln von Maschinen als „askriptives Resultat“²⁶² zu verstehen, das der Modellierung des „Denken als Hervorbringen“²⁶³ entspringt.

Beispielhaft hierfür ist auch die von Searle vorgenommene Unterscheidung zwischen dem Funktionieren und dem Verstehen.²⁶⁴ Jedoch erläutert er den Mehrwert an Verständnis, den nur der Mensch leisten kann, mit dem Intentionalitätsbegriff.²⁶⁵ In seinem Kommentar *The Chinese room revisited*²⁶⁶ (1980) macht Searle beispielsweise auf den grundlegenden Unterschied zwischen dem intentionalen Handeln des Menschen und dem „intrinsic intentional state“ einer Maschine aufmerksam. Es ist eine Unterscheidung

[...] between the conditions under which we feel justified in making attributions of intentional predicates to systems and the conditions under which the systems actually have intrinsic intentional states.²⁶⁷

Letztere existieren aufgrund von Be- und Zuschreibungen metaphorischer Art („metaphorical“, „derived and not intrinsic“²⁶⁸) und sind ferner lediglich imstande eine „formal symbol manipulation“²⁶⁹ zu leisten, während beim Menschen Intentionalität durch einen biochemischen Prozess im Gehirn angezeigt wird, durch den er einen semantischen Gehalt erfassen kann. Die Erklärung des Ursprungs der Intentionalität

²⁶⁰ Gutmann 2010, 142

²⁶¹ Hier sei auf das kantische Autonomieverständnis verwiesen.

²⁶² Gutmann 2010, 143

²⁶³ Vgl. Gutmann 2010, 144

²⁶⁴ Searle über die wesentliche Unterscheidung zwischen dem Funktionieren und dem Verstehen: „[...] the computer and the program are functioning and there is no understanding.“ (Searle 1980a, 418)

²⁶⁵ Vgl. Zwierlein 1990, 350

²⁶⁶ John R. Searle: *The Chinese Room revisited*, in: *The Behavioral and Brain Sciences*, 5 (1980b)

²⁶⁷ Searle 1980b, 345

²⁶⁸ Ebd.

²⁶⁹ Ebd.

beim Menschen mag zwar angreifbar sein²⁷⁰ – zutreffend ist jedoch die Feststellung der Diskrepanz zwischen unseren sprachlichen Be- und Zuschreibungen bestimmter Fähigkeiten und der tatsächlichen Umsetzung, wie sie auch von Weizenbaum und Gutmann angesprochen werden.

Zusammenfassend lässt sich sagen, dass sobald einer Maschine Autonomie zugeschrieben wird, dies in der Alltagssprache zunächst über die metaphorische Zuschreibung bestimmter menschlicher Fähigkeiten geschieht. Wendet man diese Ausführungen an die von Russel / Norvig vorgenommene Einteilung der Ziele der KI-Forschung in vier Herangehensweisen an, so lässt sich beurteilen, in welchem Umfang autonome Technologien tatsächlich ihrer Definition gerecht werden und in welchen Fällen weiterhin von einem vagen metaphorischen Gebrauch des Attributs autonom ausgegangen werden muss.

5.1.2. Das „Autonomie-Paradoxon“²⁷¹

Ein Aspekt ist im vorangegangenen Kapitel außen vor geblieben: Die Unmöglichkeit der Einsicht des Menschen in die Art und Weise, wie bestimmte Vorgänge einer autonomen Maschine zustande kommen, d.h. die Nachvollziehbarkeit der autonomen Aktionen, wenn eine Manipulation des Modells in einem geschlossenen autonomen System erlaubt ist. Konkret soll es nun um geschlossene autonome Systeme gehen, die nach dem *acting rationally*-Ansatz entstehen (Vgl. 3.3.2.3. (4)). Solche autonomen Artefakte werden oft als Modell biologischer Systeme beschrieben.²⁷² Dass ein System autonom ist, wird daran ersichtlich, dass es ohne Kontrolle von außen agieren kann. Es wird also in diesem speziellen Bereich der KI-Forschung zu autonomen Systemen von einem Autonomiebegriff ausgegangen, der bereits in einem anderen Fachbereich, der Biologie, modelliert wurde. Dieses naturalistische Konzept hat mit dem ursprünglich kantischen Verständnis moralischer Autonomie nichts gemein:

If Modernity thought that autonomy was a desired consequence of the human faculty of reason, contemporary naturalism aims to understand life and cognition as expressions of the autonomy of some material systems. The focus is shifted from

²⁷⁰ Siehe z. B. Dennetts Kommentar zu Searles *Chinese Room*. Daniel C. Dennett: The milk of human intentionality. Commentary/Searle: Minds, brains, programs, in: *The Behavioral and Brain Sciences* 3 (1980), 428-430

²⁷¹ In Anlehnung an Alexander Riegler: The Paradox of Autonomy: The interaction between humans and autonomous cognitive artifacts, in: Dodig Crncovic / Stuart (Hgg.): *Computation, Information, Cognition. The Nexus and the Liminal*, Newcastle 2008, 292-301

²⁷² So etwa bei George A. Bekey: „[A]utonomous systems [are] created by humans. Frequently, these systems draw inspiration from biology [...]“ (Bekey 2005, 1). Zu Autonomie als Existenzform von Systemen und Lebewesen siehe auch: Dietrich Dörner: Autonomie, in: Thomas Christaller (Hg.): *Autonome Maschinen*, Wiesbaden 2003, 112-136. Und: Alvaro Moreno / Arantza Etxeberria / Jon Umerez.: The autonomy of biological individuals and artificial models, in: *BioSystems* 91/2008, 309-319

the actions deriving from autonomy to the process able to originate the capacity, to the conditions of possibility.²⁷³

Diese Definition ist durchaus hilfreich, um künstliche autonome Systeme nach dem Vorbild des Menschen als biologisches System zu modellieren. Wie jedoch in Kapitel 4 gezeigt, kann eine naturalistische Autonomiedefinition dem Menschen nicht gerecht werden, da beispielsweise für diesen auch die Beschreibung als autonomiefähiges Vernunftwesen weiterhin gültig ist.

Doch wie ist zu beurteilen, ob ein technisches System überhaupt autonom agiert, wenn sich Autonomie lediglich am Output, nämlich der ausgeführten Aktion ablesen lässt, ein Input hierfür jedoch nicht erforderlich ist, da das System eine neue Aktion selbst generiert (etwa indem sie im Falle einer Störung „sich selbst repariert“)? Befinden wir uns in einem „Autonomie-Paradox“²⁷⁴, wenn wir nicht mehr kontrollieren können, nach welchen Gesetzmäßigkeiten ein autonomes System sich selbst kontrolliert? Gehen wir davon aus, technische autonome Systeme seien Prototypen von *Chinese Rooms*, was im Hinblick auf das letzte Zitat durchaus gerechtfertigt ist, da mit jenem naturalistischen Ansatz das Autonomieverständnis aus der körperlichen Funktionsweise eines Systems abgeleitet wird – so lässt sich das eigentümliche Verhältnis der verschiedenen Autonomiebereiche besser veranschaulichen. Mit Searle gesprochen wäre eine plausible Entgegnung, dass der als autonom bezeichneten Aktion in jedem Fall ein Input vorausgeht, in dem Sinne, dass die Maschine bereits so programmiert wird, dass im Falle einer Störung eine Lösung gefunden wird. Im Gedankenexperiment würde etwa in der Anleitung zur Aufgabenlösung stehen: „Nutze Ergebnisse aus Aufgabe *a* um ein noch unbekanntes Problem *x* in Aufgabe *b* zu lösen ohne auf Input zu warten.“ Entsprechend der Redeweise von der „schwachen“ KI ließe sich auch von der Autonomie sagen, sie sei „schwach“, da sie bloß eine partielle Nachahmung und keinesfalls eine Entsprechung der menschlichen Autonomie erbringt.

Das paradoxe Verhältnis von autonomem Mensch und autonomer Maschine ist damit aber noch nicht gelöst: Während der Mensch aufgrund metaphorischer Zuschreibung ein spezifisches Bild von der Funktionsweise der Autonomie der Maschine hat und hieraus spezifische Ziele ableitet, welche die Maschine verfolgen soll, ist die Maschine, gerade durch die ihr übertragenen Kontroll- und Autonomisierungsfähigkeiten dazu in der Lage, dem Menschen unbekannte Problemlösungen zu finden. In diesem Fall löst die Maschine eine ihr übertragene Aufgabe nicht dadurch, dass sie dieselben Methoden wie der Mensch schneller und effizienter anwendet, sondern indem sie neue Methoden findet. Riegler erläutert das Problem wie folgt:

²⁷³ Moreno et al. 2008, 310. Die Nähe zu biologischen Vorbildern wirft ebenfalls Probleme auf, wie etwa das der Definition von Leben: „Such tight a relation between life and autonomy suggests a complicated status for artificial autonomous systems. [...] if an artificial autonomous system would be produced, then we might call it “alive”.“ Ebd.

²⁷⁴ Vgl. Riegler 2008

So there are two non-overlapping domains, the actual working of the robots as e-autonomous system and the conceptual domain of the designer-user who would like the artifact to perform certain tasks in an s-autonomous way.²⁷⁵

Riegler führt in seinem Essay für technische Autonomie den Begriff „e-autonomous“ ein, während er den Menschen als „s-autonomous“ bezeichnet. Er merkt an, dass der Mensch nicht nachvollziehen kann, auf welche Art und Weise eine neue Aktion des autonomen Artefakts zustande kommt, da er über den Output lediglich die Information erhält, dass etwas geschehen ist; Mensch und Maschine setzen Autonomie unterschiedlich um. Die eigentlichen technologischen Prozesse, die unter dem Attribut „autonom“ zusammengefasst werden, sind nicht nachvollziehbar, da die Grundlagen, nach denen der Mensch seine eigene Autonomie definiert, nicht einmal nachgeahmt werden, sondern technische autonome Aktionen bestimmten mathematischen Berechnungen unterliegen. Der „s-autonomous way“ wäre in diesem Fall das aus dem menschlichen Handeln abgeleitete Modell. Der wesentliche Unterschied liegt auch hier wieder in der Unterscheidung zwischen autonom handeln und autonom sein. Die Autonomie des Menschen äußert sich sowohl in seinem Handeln als auch an seinem Sein. Dieses ist nach kantischer Auffassung nicht an der bloßen Handlung erkennbar; Autonomie ist nicht nur die Selbstbestimmung der praktischen Vernunft, sondern auch die Selbstgesetzgebung der Vernunft sowie Ausdruck von Würde und Sittlichkeit. Die Selbstgesetzgebung der Maschine hingegen tritt lediglich in Form der Selbstkontrolle zum Vorschein, um die menschenunabhängige Funktionsfähigkeit in einem bestimmten Zeitraum sicherzustellen. Dem autonomen Handeln und Sein des Menschen steht kontrollgebundenes autonomes Handeln der Maschine gegenüber. Weder der Input-Geber noch der Output-Erhaltende haben die Kontrolle über das, was innerhalb der geschlossenen autonomen Maschine geschieht, da in dieser aufgrund der gegebenen Mittel neue Erkenntnisse gewonnen werden, die von außen nicht einsehbar sind.

Auf der einen Seite ließen sich an diesem Punkt wieder jene Diskussionen über die potentielle Möglichkeit einer starken KI-Umsetzung und damit der künstlichen Entwicklung etwa von Qualia aufgreifen. Diese stelle ich jedoch zurück, denn aufgrund einer Nichtdeterminiertheit einer solchen Maschine ließe sich lediglich sicher sagen, dass sie sich in nichtvorhersehbarer Art entwickeln kann. Ob dabei tatsächlich eine dem Menschen äquivalente Form qualitativer und quantitativer Autonomie entsteht, kann ja gerade nicht vorhergesehen oder überprüft werden.

Auf der anderen Seite hat das „Autonomie-Paradox“ einen unmittelbaren Einfluss auf das Verhältnis von Kontrolle und Interaktion zwischen Mensch und Maschine. Unter Berücksichtigung dieser Tatsache können so Kriterien festgelegt werden, wie (1) die generelle Vorgabe des Zwecks der Maschine durch den Menschen, d.h. die Maschine entwickelt sich nur in dem ihr durch den Menschen vorgegebenen Rahmen, und (2) die

²⁷⁵ Riegler 2008, 298

explizite Bezeichnung eines autonomen Systems als solches sowie seines Autonomiegrads. Selbst wenn der Mensch der Maschine den Zweck vorgibt,²⁷⁶ kann das autonome System innerhalb der gesetzten Hierarchie bestimmte Aufgaben weitaus besser bewältigen als der Mensch selbst. Da dies jedoch nicht zwangsläufig ersichtlich ist, ergibt sich hieraus die Notwendigkeit, das jeweilige autonome Artefakt oder System stets als solches zu deklarieren, damit der Mensch, welcher mit einem bestimmten autonomen System interagiert weiß, um welche Art von Maschine es sich handelt. Hinsichtlich dieser Hierarchiekonstellation zwischen Mensch und Maschine muss zwangsläufig auch näher betrachtet werden, was unter dem Begriff Interaktion zu verstehen ist.

Mit der Überlegung, ob es sich nicht um Interaktion, sondern vielmehr um eine gewisse Form von Domestizierung handelt, liefert Alexander Riegler einen Einwand gegen einen Sprachgebrauch, der Gleichwertigkeit der Interaktionspartner suggerieren könnte.²⁷⁷ Dadurch dass der Mensch sich autonome Systeme für eigene Zwecke schafft und die Kommunikation mit diesen auf die Erfüllung einer bestimmten Aufgabe gerichtet ist, sei es treffender, das Verhältnis von Mensch und Maschine nicht als ein auf Interaktion beruhendes zu beschreiben, sondern mit der Domestizierung zu vergleichen.²⁷⁸ Riegler verweist auf die Tatsache, dass der Mensch sich Tieren gegenüber, welche ja mitunter als biologische autonome Systeme dargestellt werden, in einer vergleichbar paradoxen Autonomie-Situation befindet. Domestizierung von Maschinen kann zum Einen in dem Sinne verstanden werden, dass Maschinen von der Produktionshalle ausgelagert und im privaten Umfeld des Menschen Eingang finden. Die berücksichtigt jedoch nur den Aspekt der örtlichen Gebundenheit und nicht den der Zähmung, den der Domestizierungsbegriff ebenfalls birgt. Zum Anderen verweist dieser Begriff deutlich auf die Hierarchieebene, da hierdurch die Zwecksetzungsautonomie des Menschen deutlich zum Ausdruck kommt.

5.2. Autonomisierung der Technologien im Kontext von Technikgenese und humaner Selbstkonstitution

Nachdem im letzten Kapitel erörtert wurde, wie der Sprachgebrauch sich auf das Verhältnis zwischen autonomem Mensch und autonomer Maschine auswirkt, geht es nun um die bereits in Kapitel 5.1.2 teilweise angesprochene Verflechtung der autonomer Technologien in das zielorientierte Handeln des autonomen Menschen.²⁷⁹

²⁷⁶ Zur Hierarchie vom Mensch und Maschine vgl. etwa Christaller 2001, 220ff. „In den Kontexten der Robotik ist an der Zwecksetzungskompetenz von Personen grundsätzlich festzuhalten.“ (Christaller 2001, 220)

²⁷⁷ Riegler 2008, 298f

²⁷⁸ Vgl. ebd., 298

²⁷⁹ Die Perspektive einer gewissen Autonomie der Wissenschaft, bestimmte Technologien allein aufgrund ihrer Machbarkeit umzusetzen, soll an dieser Stelle nicht näher beleuchtet werden. Hierzu müsste z. B. das weite Feld der transhumanistischen Forschung betrachtet werden. Eine weitere Perspektive ist etwa

Das erste Unterkapitel befasst sich daher mit der These, dass der technische Fortschritt von Automatisierung hin zu Autonomisierung der Technik vordergründig am Nutzen für den Menschen orientiert ist und damit die Grundlage zur Entwicklung erfolgreicher Technologien. Nur so lässt sich plausibel argumentieren, weshalb die Autonomisierung technischer Systeme innerhalb des technologischen Fortschritts überhaupt notwendig ist. Im zweiten Unterkapitel soll es schließlich darum gehen, die neue Position, die sich der autonome Mensch mithilfe seiner technischen Mittel schafft, näher zu umreißen.

5.2.1. Neue Mittel, alte Zwecke

Historisch betrachtet lässt sich die These bekräftigen, dass jene Technologien, die als autonom bezeichnet werden, lediglich neue Mittel zur Erfüllung der altbekannten Zwecke Produktion und Arbeitserleichterung sind.

Um die Entwicklung von Automaten²⁸⁰ und Maschinen voranzutreiben, waren seit dem Zeitalter der Industrialisierung insbesondere folgende zusammenhängende Fragen relevant: Wie kann die Arbeit des Menschen durch technische Mittel erleichtert werden? Wie funktionieren bestimmte körperliche Arbeitsvorgänge und wie können sie technisch umgesetzt werden? Seit der Entwicklung der Kybernetik sind diese Fragen um folgende erweitert worden: Wie funktionieren die geistigen Arbeitsvorgänge des Menschen und wie können diese technisch umgesetzt werden? Ab diesem Stadium sind nun neue Zwecke als jene der Produktion denkbar.

Doch blicken wir zunächst zurück: Siegfried Giedion beschreibt in seinem Werk *Die Herrschaft der Mechanisierung*²⁸¹ jenes sich flächendeckend verbreitende Phänomen der Industrialisierung, dem das aristotelische Verständnis eines Instruments, das sich selbst bewegt, zu Grunde liegt. Anfang des 19. Jahrhunderts begann zunächst in Europa und schließlich in Nordamerika die hochentwickelte Mechanisierung der Industrie und

die Forschung zu Zwecken der Entschlüsselung der inneren Vorgänge des Menschen anhand von humanoiden Robotern. Über die verschiedenen Motivationen zur Entwicklung humanoider Roboter siehe: Michael Decker: Ein Abbild des Menschen: Humanoide Roboter, in: Michael Bölker et al. (Hgg.): Information und Menschenbild, Berlin / Heidelberg 2010, 41-62.

²⁸⁰ Bereits seit der Antike sind Automaten bekannt und gelten als solche „[...] bis in die Neuzeit hinein [als] konstruierte Statuetten, die Bewegungen und andere Funktionen von Lebewesen nachahmten zur Erregung von mythologischem oder spielerischem Staunen.“ (Helmut Schnelle: Automat, in: HWPh, Bd. 1 (A-C) 1971, 695) Der Automat dient dem Menschen jedoch nicht nur der Unterhaltung, Aristoteles beschreibt etwa in der *Politik* mit dem Adjektiv *automatos* auch jene „Instrumente, die sich von selbst bewegen und damit die Arbeit von Sklaven überflüssig machen.“ (Zit. Nach Friedrich Landwehrmann: Automation, in: HWPh, Bd. 1 (A-C) 1971, 698) Soweit sind die Zwecke eines Automaten klar definiert: zum einen als Mittel der Unterhaltung und zum anderen als Arbeitskraftersatz. Der flächendeckende Einsatz von Automaten – vordergründig als Arbeitskraftersatz, dann zur Unterhaltung – erfolgt jedoch erst seit dem Industrialisierungszeitalter.

²⁸¹ Siegfried Giedion: *Die Herrschaft der Mechanisierung*. Ein Beitrag zur anonymen Geschichte, Frankfurt a.M. 1994, Titel der Originalausgabe: *Mechanization Takes Command*, Oxford 1948

leitete damit „[...] die Eliminierung des komplizierten Handwerks [...]“²⁸² ein. Ein Grund für diese Entwicklung lag in dem Bemühen, das aller Mechanisierung zu Grunde liegende Vorbild, nämlich die menschliche Hand, für die Produktion zu optimieren.²⁸³ Erreicht wurde dies durch die maschinelle Nachahmung von Bewegungen, welche die von Menschen per Hand durchgeführten Arbeitsschritte ersetzen. Mit dem natürlichen Vorbild, der menschlichen Hand, haben diese jedoch nichts mehr gemein. Der Hand fehlt die „automatische Fertigkeit“, „sie kann nicht ununterbrochen und gleichmäßig tätig sein“²⁸⁴, wohingegen Giedion zufolge „die kontinuierlich rotierende Bewegung“²⁸⁵, die mechanische Reproduktion und das Fließband die Kennzeichen der Mechanisierung waren. Das zweite Mittel der Mechanisierung war neben dem Fließband die wissenschaftliche Betriebsführung. Diese war und ist in der Produktion von Bedeutung, um einen reibungslosen Ablauf der Produktion zu gewährleisten sowie den Ertrag zu steigern. Sie beruht auf der „[...] Untersuchung des Ablaufs der menschlichen Arbeitsleistung“²⁸⁶: Die genaue Analyse ermöglicht die Optimierung ihrer technischen Umsetzung, die Ausschaltung überflüssiger Arbeitsvorgänge und führt somit zur Rationalisierung.²⁸⁷

Indem autonome Systeme in ihrer Konstruktion eine höhere Komplexität aufweisen und der Lösung von Aufgaben mit einem hohen Schwierigkeitsgrad dienen, stellen sie gegenwärtig eine Erweiterung vorangegangener technischer Mittel dar. Im Giedion'schen Sinne ist der Fortschritt sowohl im Hinblick auf die Entwicklung von der Mechanisierung hin zur Automatisierung und schließlich zur Autonomisierung feststellbar, als auch im Bereich der wissenschaftlichen Betriebsführung, die nun in den Service- und Privatbereich übergreift. Das Bild des „synchronisierten Organismus“, bestehend aus Mensch und Maschine, das Giedion bereits Mitte des 20. Jahrhunderts für

²⁸² Giedion 1994, 22

²⁸³ Für eine ausführliche Beschreibung, wie Körper- und Bewegungsstudien Eingang in die Mechanisierung fanden, siehe Giedion 1994, 37-51.

²⁸⁴ Ebd. 70. Giedion sagt hierzu weiter: „[...] es widerspricht dem Organischen, das Wachstum und Veränderung einschließt, sich der Automatisierung zu unterwerfen.“

²⁸⁵ Ebd., 70

²⁸⁶ Ebd., 103

²⁸⁷ Die wissenschaftliche Betriebsführung (scientific management) geht auf den US-amerikanischen Ingenieur Frederick W. Taylor (1856-1915) zurück. Er beabsichtigte damit die Erleichterung der Arbeit bei gleichzeitiger Erhöhung der Leistungsfähigkeit. „Der menschliche Körper wird daraufhin untersucht, bis zu welchem Grade er in einen Mechanismus verwandelt werden kann.“ (Giedion 1994, 122). Auch die Arbeit des Betriebsingenieurs Frank B. Gilbreth (1886-1924) ist wegweisend für die wissenschaftliche Betriebsführung. Gilbreth gehört zu jenen Ingenieuren, die Bewegungsvorgänge des Menschen, z. T. auch der Natur, analysierten, um das Handwerk mechanisch umsetzen zu können – an diese entstandenen Maschinen wiederum passte Taylor die menschliche Arbeitskraft an. Dieses komplexe System der Wechselwirkung von Mensch und Maschine nennt Giedion wiederholt „synchronisierter Organismus“. Vgl. Giedion 1994, insbes. 44f, 69-76 sowie 101-126.

die frühe Zeit der Industrialisierung zeichnet, scheint den heute angestrebten autonomen Systemen umso mehr gerecht zu werden.

Erstens: Betrachtet man die von Giedion erwähnten Mittel und Zwecke der Mechanisierung, so fällt auf, dass mit der Konstruktion autonomer Systeme zwei Bereiche vereint werden sollen: (1) einen ganzheitlichen Mechanisierungsprozess in eine Maschine komprimieren und (2) in diesen menschliche Attribute zu integrieren, etwa durch den Einbau „kognitiver“ Fähigkeiten und mittels physischer Ähnlichkeit, zum Beispiel in Form von humanoiden Robotern. Hierdurch könnte das bereits von Giedion konstatierte Manko von Maschinen und Automaten, ihre Schwierigkeit in der Interaktion mit dem Menschen, der physischen Welt, insbesondere mit biologischen Materialien, behoben werden. Innerhalb der Entwicklungslinie des Fortschritts der Mechanisierung bilden autonome Systeme damit eine konsequente Fortführung der Technisierung. Längst existieren in der industriellen Produktion „autonomous work groups“²⁸⁸, in denen Mensch und Maschine zusammenarbeiten. In einem solchen Modell liegt die Aufgabe des Arbeiters darin, die Qualität und Quantität der Produktion zu überwachen und zu erhöhen, wobei der Erfolg die Bezahlung beeinflusst. Dies ist der ideelle Mehrwert der Arbeitsleistung des Menschen, während rein funktionale Aufgaben autonomen Teilsystemen und Robotern übertragen werden.²⁸⁹ Anzumerken ist jedoch, dass innerhalb der Autonomisierung der Technologien zwischen den einzelnen Zwecken, für die jene Techniken zum Einsatz kommen, unterschieden werden muss. Sie sind nicht länger allein auf den industriellen Sektor beschränkt, so wie es zu Zeiten der Mechanisierung war, sondern greifen auf den Service-Bereich über.²⁹⁰ Mechanische und automatisierte Fertigungsprozesse wurden aus dem menschlichen Lebensraum in eine Produktionshalle ausgelagert, autonome Systeme und Serviceroboter sollen hingegen gezielt an den menschlichen Lebensraum angepasst werden.

Zweitens: Noch weitgehend unbeachtet geblieben ist bis heute der Aspekt, dass mit der zunehmenden Technisierung einhergegangene wissenschaftliche Betriebsführung ebenfalls weiterentwickelt wurde und für den aktuellen Stand der betreffenden Technologien mitverantwortlich ist. Dabei liefert dies Rückschlüsse auf die Ziele und Motivationen des Fortschritts in Richtung Autonomisierung. Ich möchte an dieser Stelle anstatt des Attributs „autonom“ für jene Systeme vielmehr den betriebswirtschaftlich motivierten technischen Fortschritt hervorheben: Das Ziel, die „Steigerung der mechanischen Leistungsfähigkeit“²⁹¹ und damit die Erhöhung der Produktion, erfordert die Anpassung der menschlichen Arbeitskraft an den mechanischen Prozess, sofern die

²⁸⁸ Roger L. Anderson / William B. Gartner: When Robots and People work together, in: *Robotics 1* (1985), 69

²⁸⁹ Vgl. Anderson / Gartner 1985, 69ff

²⁹⁰ Vgl. Thrun 2004, 11

²⁹¹ Giedion 1994, 124

Maschine bestimmte Aufgaben nicht ausführen kann.²⁹² Im Zeitalter der Vollmechanisierung funktioniert der Fertigungsprozess weitgehend automatisch – die Aufgabe des Menschen liegt darin, die Maschinen und Automaten zu überwachen und korrigierend einzugreifen, seine Handgriffe sind an den mechanischen Prozess angepasst. Sobald jedoch selbst diese Aufgaben von den Maschinen übernommen werden können, ist der Übergang von der Automatisierung zur Autonomisierung des Produktionssystems markiert. Dies ist ein klassisches Beispiel für geschlossene autonome Systeme, in denen, wie in Kapitel 3.3.2.3. näher ausgeführt, Kontrolle und KI die wesentlichen Kriterien der Autonomie sind. Der Fortschritt in der wissenschaftlichen Betriebsführung wird daran sichtbar, je mehr ein Fertigungsprozess ohne den Eingriff durch den Menschen funktioniert.

Im industriellen Bereich ist die Autonomie der Maschine dadurch, dass sie durch menschliche Ziele determiniert ist, klar definierbar. Dies gilt ebenfalls für den tertiären Sektor, sofern diese durch die Rationalisierung bestimmter Arbeitsprozesse motiviert ist, eine Arbeitserleichterung darstellt oder der Unterhaltung dienen soll, etwa im Falle der Thrun'schen *professional service robots* oder *personal service robots*. Der Zweck der Maschine ist nicht mehr die Produktion, sondern die Dienstleistung.

5.2.2. Neue Mittel, neue Zwecke?

Wie sehen die Erweiterungen der Autonomie des Menschen aus, wenn offene oder geschlossene autonome Systeme in den menschlichen Lebensraum einbezogen werden? Diese Entwicklung scheint zunächst zweischneidig, zumal mit der Entstehung eine zunehmende Abhängigkeit des Menschen von seinen selbstgeschaffenen autonomen Systemen einhergeht.

Es ist jedoch zu unterscheiden zwischen der Autonomiegestaltung des Menschen selbst sowie den Autonomiemöglichkeiten der Maschine. Bislang hat sich gezeigt, dass die Definition von Autonomie in der Entwicklung von autonomen technischen Systemen und Artefakten zwar einem gewissen biologischen Modell von Autonomie folgt, was auch für den Menschen als biologisches Wesen zutrifft. Von diesem Stand der Entwicklung sind die Autonomiemöglichkeiten der Maschine überschaubar, wenn auch nicht im Detail nachvollziehbar. Die Überschaubarkeit resultiert aus der Zwecksetzungsautonomie des Designers und Users. Es gibt demzufolge jedoch keine haltbaren Ansatzpunkte, die die Notwendigkeit belegen, autonome artifizielle Systeme zu produzieren, die sich der Zwecksetzungsautonomie des Menschen entledigen könnten. Hier sei an ein Zitat Dennetts erinnert:

²⁹² Vgl. Giedion 1994, 124 über das Verhältnis von Mensch und Maschine: „In der gegenwärtigen Situation, in der die Maschine nicht weit genug entwickelt ist, um gewisse Operationen auszuführen, verlangt der Taylorismus dagegen von der Masse der Arbeiter nicht Initiative, sondern Automatisierung. Die menschlichen Bewegungen werden zu Hebeln der Maschine.“

Yes, I do think that it's possible to program self-consciousness into a computer. Self-consciousness can mean many things. [...] As we develop good accounts of consciousness, it will no longer seem so obvious to everyone that the idea of a self-conscious is a contradiction in terms. At the same time, I doubt that there will ever be self-conscious robots. But for boring reasons. There won't be any point in making them.²⁹³

Dennett macht darauf aufmerksam, dass es für einen Roboter mit Selbstbewusstsein überhaupt keinen Zweck gibt; die Kosten für einen solchen Roboter stünden in keiner Relation zu seinem nicht vorhandenen Nutzen für den Menschen.²⁹⁴ Auch wenn sich Dennett in dieser Passage zwar nicht auf Autonomie, sondern auf das Bewusstsein bezieht, so macht er auf etwas Entscheidendes aufmerksam: auf die Tatsache, dass wir nicht nur Maschinen Fähigkeiten zuschreiben, sondern in erster Linie uns selbst. Somit hält er es für nicht ausschließbar, etwa Bewusstsein technisch umzusetzen – es kommt lediglich auf die Prämissen an, nach denen wir Bewusstsein als solches gelten lassen. Dennett bezeichnet solche Überzeugungen über exklusiv dem Menschen zustehende Fähigkeiten wie Bewusstsein, aber auch Autonomie, als „illusion of common sense“.²⁹⁵ Autonomie ist, nach bestimmten Prämissen, graduell auch Artefakten zuzuschreiben. Diese Prämissen sind folglich solche, die Autonomie als mathematisch zu berechnendes, teilweise auch körperlich umzusetzendes Finden von Handlungsoptionen definieren. Dem Menschen steht jedoch weiterhin frei, sich selbst auf einer anderen Ebene, etwa einer moralischen, als autonom zu begreifen oder diese wie Dennett als „commonsense illusion“²⁹⁶ abzutun.

Der Mensch definiert seine Fähigkeiten mittels seiner neuen Technologien jedoch selbst neu. Mithilfe von bestimmten autonomen Systemen eröffnet sich der Mensch neue Handlungsoptionen, die ohne solche Technologien undenkbar gewesen wären. Mit dem Zugewinn an Handlungsmöglichkeiten durch autonome Systeme und damit an quantitativer Autonomie geht jedoch gleichzeitig der allmähliche Verlust jener Kompetenzen einher, die der Mensch der Maschine überträgt.

Hannah Arendt (1906-1975) beschreibt in der *Vita Activa* diesen Zustand, in dem die Grenze zwischen Mensch und Technik zunehmend verwischt, ebenfalls als einen von sowohl positiven als auch negativen Auswirkungen geprägten:

Im Experiment kann die hypothetisch vorweggenommene Welt jederzeit zu einer wirklichen, von Menschen verwirklichten Welt [werden], was zwar sagt, dass die praktischen Vermögen des Menschen, das Vermögen zu handeln, herzustellen, ja sogar Welten zu erschaffen, unvergleichlich größer und mächtiger sind, als irgendein vergangenes Zeitalter zu träumen wagte; was aber andererseits auch

²⁹³ Daniel C. Dennett: Can Machines Think? Postscript [1985]: Eyes, Ears, Hands and History, in: Christof Teuscher (Hg.): Alan Turing: Life and Legacy of a Great Thinker, Berlin / Heidelberg 2004, 310f

²⁹⁴ Vgl. Dennett 2004, 311

²⁹⁵ Ebd., 311

²⁹⁶ Ebd.

heißt, dass die volle Ausnutzung gerade seines welterschaffenden Vermögens den Menschen in das Gefängnis seiner selbst, seines eigenen Denkvermögens verweist, ihn unerbittlich auf sich zurückwirft, ihn gleichsam in die Grenzen seiner selbst geschaffenen Systeme sperrt.²⁹⁷

Arendt zufolge hat der Mensch durch die Zuhilfenahme moderner Technologien sämtliche Arbeitsprozesse vereinfacht. Damit einhergehend wird jedoch die Grenze zwischen dem Menschen als Naturwesen und seiner gemachten, aus Maschinen bestehenden Welt aufgehoben.²⁹⁸ Die volle Ausnutzung aller ermöglichbaren Technologien bedeutet für Arendt gleichzeitig ein Gefangensein in der Ausdehnung der technisierten Welt, aus der kein Weg in die Natur zurückführt. Aufgrund der Tatsache, dass durch Zuhilfenahme von Techniken die Lebensumstände des Menschen verbessert werden, muss der technologische Fortschritt daher nicht zwangsläufig als eine Verknechtung des Menschen durch seine selbst geschaffene Maschine aufgefasst werden. Das von Arendt konstatierte Zurückgeworfensein auf sich selbst kann ebenfalls als Chance aufgefasst werden, sich selbst als Mensch neu zu definieren, so wie es in neuen Konzeptionen von Menschenbildern zunehmend der Fall ist.

Man denke etwa an Donna Harraways Manifest: „[...] wir sind alle Cyborgs“²⁹⁹, oder Nicole Karafyllis' Neologismus „Biofakt“ für bio- oder gentechnisch manipulierte Wesen, die dennoch dem natürlichen Wachstum unterliegen.³⁰⁰ In ihrem Manifest schreibt Harraway über die Grenzen zwischen Mensch und Technik:

Dieses Essay ist ein Plädoyer dafür, die Verwischung dieser Grenzen zu genießen und Verantwortung bei ihrer Konstruktion zu übernehmen.³⁰¹

Ebendiese Konstruktionen, die der autonome Mensch dank des technischen Fortschritts bauen kann, untersucht wiederum Karafyllis und gibt ihnen den Namen „Biofakte“. Mit dem Neologismus „Biofakt“ – eine Verknüpfung aus den Begriffen „Bios“ und „Artefakt“ sowie deren Merkmalen – sind jene Hybride gemeint, die weder Technik noch Mensch oder Natur zugeordnet werden können, da sie eine Schnittstelle von beiden bilden:

²⁹⁷ Hannah Arendt: *Vita activa*, München 42006, 365f

²⁹⁸ Vgl. Kohji Ishihara: *Werkzeuge, Maschinen und menschliche Autonomie*, in: Holger Burkhardt / Jürgen Sikora (Hgg.): *Praktische Philosophie in gesellschaftlicher Perspektive. Ein interdisziplinärer Diskurs*, Münster 2005, 79f

²⁹⁹ Donna Harraway: *Ein Manifest für Cyborgs. Feminismus im Streit mit den Technowissenschaften*. In: Claus Pias u.a. (Hgg.): *Kursbuch Medienkultur, Die maßgeblichen Theorien von Brecht bis Baudrillard*, Düsseldorf 2004, 465. Cyborg (Akronym, cybernetic organism) ist die Bezeichnung für ein Mischwesen aus Maschine und lebendigem Organismus.

³⁰⁰ Siehe Nicole C. Karafyllis: *Biofakte - Grundlagen, Probleme, Perspektiven*, in: *Erwägen Wissen Ethik* 17/4 (2006), 547-558

³⁰¹ Harraway 2004, 465

Biofakte problematisieren *in corpore* einen Wachstums- und Bewegungsbegriff, der bislang als Unterscheidungsmerkmal zwischen den Sphären der Natur und Technik gedient hat.³⁰²

Nach Karafyllis ist das Wachstum ein wesentliches Unterscheidungsmerkmal zwischen Natürlichem und Künstlichem; der Begriff Biofakt soll diesem Umstand Rechnung tragen. Ein weiteres Beispiel, wie sich der autonome Mensch mittels der Technologien, die ihn umgeben, neu definieren kann, ist Bruno Gransches Konzept, nach dem der Mensch sich selbst als Autofakt auffasst, das sich in seiner selbst geschaffenen „New Reality“ bewegt:

Die New Reality ist eine Sphäre, in der Technik und Organismus, Virtualität, Simulation und Realität miteinander verschmelzen. Das Neue ist keine der Komponenten für sich, sondern die brisante Kombination der Teile.³⁰³

New Reality ist somit ein Synonym für die „selbst geschaffenen Systeme“³⁰⁴, doch ist dieser Raum nicht wie bei Arendt als Gefängnis, sondern als Raum für vielfältige Gestaltungsmöglichkeiten des Menschen zu verstehen. Der Begriff Autofakt rekuriert nun auf das subjektive Vermögen des Menschen, sich innerhalb der technischen Welt selbst als technisches Wesen zu begreifen und zu gestalten, denn „[...] der Mensch als Autofakt ist nicht nur Autokonstrukt, ein Produkt des Selbstdesigns, sondern auch selbst Schaffender und ein sich schaffender Selbstgestalter.“³⁰⁵ Die Anforderungen an das Autofakt ähneln somit denen der Harraway'schen Cyborgs.

Der Rahmen, in welchem der Mensch in den oben beschriebenen Beispielen jeweils handelt, bietet für die Definition von Autonomie einen neuen Status, wenn nicht sogar eine neue Funktion. Setzt man ein Selbstverständnis des Menschen voraus, das seine Technisierung nicht ausschließt, so können die Aussagen über die Autonomie des Menschen neu strukturiert werden:

(1) Die Autonomie der Maschine ist stets an die Zwecksetzungsautonomie des Menschen gekoppelt, da dieser ein bestimmtes Ziel verfolgt. Dieses ist in einem neuen Menschenbild beispielsweise mit der Selbstverwirklichung des Menschen als Cyborg oder als Autofakt beschreibbar, oder in der Produktion von Biofakten.

(2) Die Quantität an autonomen Handlungsmöglichkeiten des Menschen steigt, da auch autonome Aktionen von technologischen Teilsystemen, mit denen der Mensch interagiert, von der Zwecksetzungsautonomie des Menschen geleitet sind. Autonome Maschinen und Teilsysteme sind als Ausführende bestimmter Handlungen des

³⁰² Nicole C. Karafyllis: Biofakte als neue Kategorie der Informatik? In: Raimund Jakob u.a. (Hg.): Auf dem Weg zur Idee der Gerechtigkeit. Gedenkschrift für Ilmar Tamello, Wien / Berlin 2009, 251

³⁰³ Bruno Gransche: Der Mensch als Autofakt. Technik-Haben und Technik-Sein in der New Reality des 21. Jahrhunderts, Saarbrücken 2010, 8. Ein anderes Beispiel ist das Biofakt.

³⁰⁴ Arendt 2006, 365

³⁰⁵ Gransche 2010, 6

Menschen anzusehen. Der Mensch selbst eröffnet sich hingegen neue Dimensionen als Handlungsraum.

(3) Der Verlust an Autonomie in Form von Übertragung autonomer Handlungsmöglichkeiten auf bestimmte technische Hilfssysteme wird vor allem durch die Kontrollmöglichkeit kompensiert. Übernimmt eine autonome Maschine die Aufgabe des Menschen, so tritt Kontrolle an die Stelle von Autonomie. Kontrolle ist zum einen zu verstehen als jene den Technologien implementierte Fähigkeit zur Selbstkontrolle und zum anderen die Kontrollmöglichkeit durch den Menschen von außen. In diesem Fall stellt sich die Frage nach der Zuordnung von Verantwortlichkeit.³⁰⁶ Hieran zeigt sich, dass der vermeintliche Kontrollverlust gegenüber der Maschine entweder durch mangelndes Verständnis begründet ist oder aber dass die Kontrollmöglichkeit bei einem anderen Menschen liegt, welcher damit nicht nur die Kontrollmacht über die Technologie, sondern auch über deren Nutzer hat.

³⁰⁶ Vgl. Christaller 2001, 36

6. Schlussbemerkung

Die Betrachtung der unterschiedlichen Autonomiedefinitionen erfolgte unter Berücksichtigung der Frage, wie sich die wachsende Kooperation und Interaktion von Mensch und Maschine auf unser Verständnis von Autonomie auswirkt.

Im Verlauf dieser Arbeit haben sich diesbezüglich folgende drei Punkte ergeben: (1) Der technische Autonomiebegriff unterliegt einer eigenständigen, von der Philosophie unabhängigen Definition. (2) Moralphilosophisch betrachtet wird die Autonomie des Menschen von autonomen Technologien weder bereichert noch angegriffen. (3) Im Rahmen einer epistemisch-teleologischen Handlungsmotivation wird die Autonomie des Menschen durch technische Autonomie erweitert, indem sich der Mensch mittels autonomer Technologien den eigenen Handlungsraum sowie seine Handlungsoptionen erweitert. So entstand in Kapitel fünf der Vorschlag, das Konzept der Zwecksetzungsautonomie mit der Unterscheidung von quantitativer und qualitativer Autonomie zu erweitern, wobei quantitative Autonomie als Vielfalt der Handlungsmöglichkeiten und qualitative Autonomie in moralischer Hinsicht als begriffliche Abgrenzung zur quantitativen Autonomie verstanden werden soll.

Die Zuhilfenahme bestimmter Techniken bewirkt grundsätzlich eine Veränderung des Handlungsraums des Menschen und so beeinflussen jene autonomen Technologien, wie alle Technologien vor ihnen, ebenfalls die Umwelt des Menschen. Der erläuterte quantitative Zugewinn der Autonomie des Menschen entspringt nun den selbstgeschaffenen, neuen Umweltbedingungen des pragmatisch-zielorientierten autonomen Handelns des Menschen, durch welches die Vernetzung von Mensch und Technik weiter ansteigt. Gleichzeitig ist der Mensch in einem kantischen Verständnis unter dem Gesichtspunkt der Selbstesetzgebung des freien Willens autonom, ohne dass bestimmte Technologien hierauf Einfluss hätten. Darüber hinaus muss jedoch entschieden werden, wie viele seiner Fähigkeiten der Mensch durch technische Hilfsmittel ersetzen und optimieren will; dies wirft jedoch Fragen ethischer Art auf.

Sieht man die Zukunft des Menschen als mensch-technisches Hybridwesen, so eröffnen sich dem Menschen auf der einen Seite vollkommen neue Gestaltungsmöglichkeiten, indem der Mensch bestimmte Fähigkeiten an technische Teilsysteme abgibt, die diese effektiver beherrschen. Jene können zwar nur solche sein, die auch technisch realisierbar sind, doch setzt die Realisierbarkeit stets eine neue Erörterung dessen voraus, was unter einer bestimmten Fähigkeit verstanden werden soll. Kategorisch ließe sich so nicht einmal mehr die künstliche Nachbildung des Bewusstseins ausschließen, man denke etwa an die Argumentationsweise Dennetts. Wie gezeigt, hängt die Autonomie von Maschinen von ihrer Beschreibung des Menschen ab und muss demnach stets als solche gekennzeichnet sein; kantianisch gesprochen hieße dies: die Maschine ist als Mittel zu einem Zweck autonom und nicht als Zweck an sich selbst. Die Gültigkeit dieser Aussage ging auch den Ausführungen zum historischen Hintergrund von der Automatisierung hin zur Autonomisierung hervor; auch wurde hier das

unveränderte Verhältnis von Mensch und Maschine deutlich, wonach die Fähigkeiten des Menschen stets als Vorbild dienen. Dem Menschen kommt Autonomie zu, weil er Mensch ist, der Maschine kommt sie erst dann zu, wenn der Mensch sie ihr zugeschrieben hat und sie derart gebaut hat, dass sie die Kriterien, nach welchen die Zuschreibung erfolgte, erfüllt. Zur gezielteren Abgrenzung zwischen den unterschiedlichen Formen autonomer Technik habe ich daher die Unterscheidung zwischen offenen und geschlossenen autonomen Systemen eingeführt.

Differenziert man auf diese Weise den semantischen Gehalt des Autonomiebegriffs an der Schnittstelle von Mensch und Maschine, sollte es auch nicht zu Missverständnissen führen, wenn etwa Thrun in seinem Paper argumentiert, ein Roboter müsse autonomer sein, um besser mit dem Menschen interagieren zu können. Es ist der Mensch, der sowohl seine eigene Autonomie, als auch die der Maschine erfunden hat. Es liegt daher in seiner Verantwortung stets für ein klares Verständnis zu sorgen, was im Laufe des technischen Fortschritts unter diesem Begriff aufgefasst werden soll.

Literaturverzeichnis

- Anderson, Roger L. / William B. Gartner: When Robots and People work together, in: *Robotics* 1 (1985), 69-76
- Arendt, Hannah: *Vita activa*, München ⁴2006, Titel der amerikan. Originalausgabe: *The Human Condition*, Chicago 1958
- Aristoteles: *Nikomachische Ethik*, auf d. Grundlage der Übers. von Eugen Rolfes, hrsg. von Günther Bien, Hamburg 1985
- Backus, John: Can Programming Be Liberated from the von Neumann Style? A Functional Style and Its Algebra of Programs, in: *Communications of the ACM* 21/8 (1978), 613-641
- Bagnell, Andrew / David Bradley et al.: Learning for Autonomous Navigation. Advances in Machine Learning for Rough Terrain Mobility, in: *IEEE R&A* 17/2 (2010), 74-84
- Bennett, Maxwell / Daniel Dennett / Peter Hacker / John Searle: *Neurowissenschaft und Philosophie. Gehirn, Geist und Sprache*, aus dem Engl. von Joachim Schulte, Berlin 2010 (Titel der amerikan. Originalausgabe: *Neuroscience and Philosophy. Brain, Mind, and Language*, New York 2007)
- Bensalem, Saddek / Mathieu Gallien et al.: Designing Autonomous Robots. Toward a More Dependable Software Architecture, in: *IEEE R&A* 16/1 (2009), 67-77
- Bekey, George A.: *Autonomous Robots. From Biological Inspiration to Implementation and Control*, Cambridge (Massachusetts) 2005
- Bertschinger, Nils et al.: Autonomy: An information theoretic perspective, in: *BioSystems* 91 (2008), 331-345
- Birk, Andreas / Narunas Vaskevicius et al.: 3-D Perception and Modeling. Motion-Level Teleoperation and Intelligent Autonomous Functions, in: *IEEE R&A* 16/4 (2009), 53-60
- Cheng, Gordon / Alexander Zelinsky: Supervised Autonomy: A Framework for Human-Robot Systems Development, in: *Autonomous Robots* 10 (2001), 251-266
- Christaller, Thomas et al.: *Robotik. Perspektiven für menschliches Handeln in der zukünftigen Gesellschaft*, Berlin / Heidelberg 2001
- Christaller, Thomas / Josef Wehner: *Autonomie der Maschinen – Einführung in die Diskussion*, in: Thomas Christaller (Hg.): *Autonome Maschinen*, Wiesbaden 2003, 9-35
- Cohen, Hermann: *Ethik des reinen Willens*, Nachdr. der 2. revidierten Aufl. Berlin 1907, Hildesheim 1981

- Cox Brian R. / Jeffrey L. Krichmar: Neuromodulation as a Robot controller. A Brain-Inspired Strategy for Controlling Autonomous Robots, in: *IEEE R&A* 17/1 (2010), 72-80
- Decker, Michael: Ein Abbild des Menschen: Humanoide Roboter, in: Michael Bölker et al. (Hgg.): *Information und Menschenbild*, Berlin / Heidelberg 2010, 41-62
- Dennett, Daniel C.: The milk of human intentionality. Commentary/Searle: Minds, brains, programs, in: *The Behavioral and Brain Sciences* 3 (1980), 428-430
- Dennett, Daniel C.: Can Machines Think? Postscript [1985]: Eyes, Ears, Hands and History, in: Christof Teuscher (Hg.): *Alan Turing: Life and Legacy of a Great Thinker*, Berlin / Heidelberg 2004, 310f
- Dörner, Dietrich: Autonomie, in: Thomas Christaller (Hg.): *Autonome Maschinen*, Wiesbaden 2003, 112-136
- Eisler, Rudolf: *Kant-Lexikon. Nachschlagewerk zu Kants sämtlichen Schriften / Briefen und handschriftlichem Nachlass*, Nachdruck der Ausgabe Berlin 1930, Hildesheim 2008
- Feil-Seifer, David / Maja J. Mataric: Socially Assistive Robotics. Ethical Issues Related to Technology, in: *IEEE R&A* 18/1 (2011), 24-31
- Fichte, Johann Gottlieb: *System der Sittenlehre nach den Principien der Wissenschaftslehre* (1798), in: I. H. Fichte (Hg.): *Fichtes Werke*, Bd. IV: *Zur Rechts- und Sittenlehre*, (Nachdruck der Ausgabe Berlin 1845/46) Berlin 1971
- Forschner, Maximilian: *Gesetz und Freiheit. Zum Problem der Autonomie bei I. Kant*, München/Salzburg 1974
- Giedion, Sigfried: *Die Herrschaft der Mechanisierung. Ein Beitrag zur anonymen Geschichte*, m. e. Vorwort hrsg. von Henning Ritter, m. e. Nachwort von Sanislaus von Moos, Frankfurt a. M. 1994, Titel der Originalausgabe: *Mechanization Takes Command*, Oxford 1948
- Goldammer, Eberhard von / Jochen Paul: *Autonomie in Biologie und Technik. Kognitive Netzwerke – Artificial Life – Robotik*, in: Axel Ziemke / Rudolf Kaehr (Hgg.): *Realitäten und Rationalitäten*, Berlin 1996, 277-298
- Gransche, Bruno: *Der Mensch als Autofakt. Technik-Haben und Technik-Sein in der New Reality des 21. Jahrhunderts*, Saarbrücken 2010
- Greve, Jens: *Jürgen Habermas, Eine Einführung*, Konstanz 2009
- Grunwald, Armin / Yannick Julliard: Technik als Reflexionsbegriff. Überlegungen zur semantischen Struktur des Redens über Technik, in: *Philosophia naturalis* 42/1 (2005), 127-157
- Gutmann, Mathias: *Autonome Systeme und der Mensch: Zum Problem der medialen Selbstkonstitution*, in: Stefan Selke / Ulrich Dittler (Hgg.): *Postmediale*

- Wirklichkeiten aus interdisziplinärer Perspektive. Weitere Beiträge zur Zukunft der Medien, Hannover 2010, 127-148
- Gutmann, Mathias / Benjamin Rathgeber: Zur Leistung und Funktion von Metaphern, in: Michael Bölker et al. (Hgg.): Menschenbilder und Metaphern im Informationszeitalter, Berlin 2010, 13-43
- Gutmann, Mathias / Benjamin Rathgeber / Tareq Syed: Autonome Systeme und evolutionäre Robotik: neues Paradigma oder Missverständnis?, in: Mathias Maring (Hg.): Fallstudien zur Ethik in Wissenschaft, Wirtschaft, Technik und Gesellschaft, Karlsruhe 2011, 185-197
- Habermas, Jürgen: Theorie des kommunikativen Handelns, Bd. 1 Handlungsrationalität und gesellschaftliche Rationalisierung, Frankfurt a. M. ⁴1987
- Habermas, Jürgen: Faktizität und Geltung, Frankfurt a. M. 1992
- Haraway, Donna (1985): Ein Manifest für Cyborgs. Feminismus im Streit mit den Technowissenschaften. In: Claus Pias u.a. (Hgg.): Kusrbuch Medienkultur, Die maßgeblichen Theorien von Brecht bis Baudrillard, Düsseldorf 2004, 464-471, Titel der Originalausgabe: Manifesto for Cyborgs: Science, Technology, and Socialist Feminism in the 1980's, in: *Socialist Review* 80 (1985) 65-108
- Haun, Matthias: Handbuch Robotik, Berlin / Heidelberg 2007
- Henrich, Dieter: Selbstverhältnisse, Stuttgart 1993
- Horn, Christoph: Antike Lebenskunst. Glück und Moral von Sokrates bis zu den Neuplatonikern, München 1998
- Höffe, Otfried: Ausblick: Aristoteles oder Kant – wider eine plane Alternative, in: ders. (Hg.): Aristoteles: Die Nikomachische Ethik, Berlin 1995, 277-304
- IEEE Robotics & Automation* 16/1 (2009)
- IEEE Robotics & Automation* 16/3 (2009)
- IEEE Robotics & Automation* 16/4 (2009)
- IEEE Robotics & Automation* 17/1 (2010)
- IEEE Robotics & Automation* 17/2 (2010)
- IEEE Robotics & Automation* 17/4 (2010)
- IEEE Robotics & Automation* 18/1 (2011)
- Ishihara, Kohji: Werkzeuge, Maschinen und menschliche Autonomie, in: Holger Burkhardt / Jürgen Sikora (Hgg.): Praktische Philosophie in gesellschaftlicher Perspektive. Ein interdisziplinärer Diskurs, Münster 2005, 75-86
- Janich, Peter: Was ist der Mensch?, in: Detlev Ganten et al. (Hgg.): Was ist der Mensch? Berlin 2008, 124-126
- Kant, Immanuel: Kritik der reinen Vernunft (KrV), nach der 1. und 2. Originalausgabe, hrsg. von Jens Timmermann, Hamburg 1998

- Kant, Immanuel: Kritik der praktischen Vernunft (KpV), mit einer Einleitung, Sachanmerkungen und einer Bibliographie von Heiner F. Klemme, hrsg. von Horst D. Brandt und Heiner F. Klemme, Hamburg 2003
- Kant, Immanuel: Kritik der Urteilskraft (KU), mit Einleitungen und Bibliographie, hrsg. von Heiner F. Klemme, Hamburg 2009
- Kant, Immanuel: Metaphysik der Sitten (MdS), unveränderter Abdruck 1966 der 4. Auflage 1922, hrsg. von Karl Vorländer, Hamburg 1922
- Kant, Immanuel: Grundlegung zur Metaphysik der Sitten, mit einer Einleitung, hrsg. von Bernd Kraft und Dieter Schönecker, Hamburg 1999
- Karafyllis, Nicole C.: Biofakte - Grundlagen, Probleme, Perspektiven, in: *Erwägen Wissen Ethik* 17/4 (2006), 547-558
- Karafyllis, Nicole C.: Biofakte als neue Kategorie der Informatik? In: Raimund Jakob u.a. (Hg.): Auf dem Weg zur Idee der Gerechtigkeit. Gedenkschrift für Ilmar Tamello, Wien / Berlin 2009, 249-262
- Klauser, Theodor (Hg.): Reallexikon für Antike und Christentum : Sachwörterbuch zur Auseinandersetzung des Christentums mit der antiken Welt, Bd I.: A und O – Bauen, Stuttgart 1950
- Knasek, T. M.: Robotics and Autonomous Systems. Editorial, in: *Robotics* 4 (1988), 1-2
- Kreimendahl, Lothar: Einleitung. Die Philosophie des 18. Jahrhunderts als Philosophie der Aufklärung, in: ders. (Hg.): *Philosophen des 18. Jahrhunderts*, Darmstadt 2000, 1-32
- Lakoff, George / Mark Johnson: *Philosophy in the Flesh. The Embodied Mind and Its Challenge to Western Thought*, New York 1999
- Landwehrmann, Friedrich: Automation, in: *HWPPh*, Bd. 1 (A-C) 1971, 698f
- Lämmel, Uwe / Jürgen Cleve: *Künstliche Intelligenz*, München³2008
- Lenk, Hans: *Das flexible Vielfachwesen: Einführung in die moderne philosophische Anthropologie zwischen Bio-, Techno- und Kulturwissenschaften*, Weilerswist 2010
- Luhmann, Niklas: *Zweckbegriff und Systemrationalität*, Tübingen 1977
- Mainzer, Klaus: *Leben als Maschine? Von der Systembiologie zur Robotik und Künstlichen Intelligenz*, Paderborn 2010
- Marani, Giacomo / Song K. Choi: Underwater Target Localization. Autonomous Intervention with the DIDSON sonar in SAUVIM, in: *IEEE R&A* 17/1 (2010), 64-70
- Moreno, Alvaro / Arantza Etxeberria / Jon Umerez: The autonomy of biological individuals and artificial models, in: *BioSystems* 91/2008, 309-319
- Müller, Jörn: Glück als Vollendung menschlicher Natur. Die eudaimonistische Tugendethik des Aristoteles, in: Hanns-Gregor Nissing / Jörn Müller (Hgg.):

- Grundpositionen philosophischer Ethik. Von Aristoteles bis Jürgen Habermas, Darmstadt 2009, 23-53
- Pieper, Annemarie: Autonomie, in Wilhelm Korff / Lutwin Beck / Paul Mikat (Hgg.): Lexikon der Bioethik Bd. 1, Gütersloh 1998, 289-293
- Pieper, Annemarie: Einführung in die Ethik, Tübingen / Basel ⁴2000
- Piepmeyer, Rainer: Intelligenz, *Intelligensia, Intellektueller*, in: HWPh Bd. 4 (I-K) 1976, 445-447
- Pinsky, Leonard: Do machines think about machines thinking? In: *Mind* 60 (1951) 397-398
- Pohlmann, Rosemarie: Autonomie, in: HWPh Bd. 1 (A-C) 1971, Sp. 701-719
- Poller, Horst: Die Philosophen und ihre Kerngedanken, Freiburg 2007
- Riegler, Alexander: The Paradox of Autonomy: The interaction between humans and autonomous cognitive artifacts, in: Dodig Crncovic / Stuart (Hgg.): *Computation, Information, Cognition. The Nexus and the Liminal*, Newcastle 2008, 292-301
- Richards, Jay W. (Hg.): *Are we spiritual machines?: Ray Kurzweil vs. the critics of strong AI*, Seattle/Washington 2002
- Robotics 3 (1987), Conference Reports, Intelligent Autonomous Systems
- Robotics 4 (1988)
- Rohls, Jan: *Geschichte der Ethik*, Tübingen 1991
- Rohrmoser, Günter: Autonomie, in: Herman Krings / Hans Michael Baumgartner / Christoph Wild (Hgg.): *Handbuch philosophischer Grundbegriffe*, Bd. 1 *Das Absolute – Gesellschaft*, München 1973, 155-170
- Röttgers, Kurt: Intelligenz; *Intellektueller II. V (soz.)*, in: HWPh Bd. 4 (I-K) 1976, 447f. u. 454-458
- Russell, Stuart J. / Peter Norvig: *Artificial Intelligence. A Modern Approach*, Boston / München u. a. ³2010
- Sandel, Michael J.: *Plädoyer gegen die Perfektion*, aus dem amerikan. von Rudolf Teuwsen, Berlin 2008
- Saygin, Ayse Pinar / Ilyas Cicekli / Varol Akman: Turing Test: 50 Years Later, in: *Minds and Machines* 10 (2000), 463-518
- Schanze, Jens: *Plug & Pray*, Dokumentarfilm, Mascha Film Judith Malek-Mahdavi & Jens Schanze GbR 2009
- Scherer, Georg: Person III 1-9, in: HWPh Bd. 7 (P-Q) 1989, 300-319
- Schnelle, Helmut: Automat, in: HWPh Bd. 1 (A-C) 1971, 695-697
- Searle, John R.: Minds, brains, and programs, in: *The Behavioral and Brain Sciences*, 3 (1980a), 417-457
- Searle, John R.: The Chinese Room revisited, in: *The Behavioral and Brain Sciences*, 5 (1980b), 345-348

- Searle, John R.: *Minds, brains and science (The 1984 Reith Lectures)*, Cambridge, Massachusetts ¹³2003
- Seo, Jae-sun / Bernard Brezzo / Yong Liu et al.: A 45nm CMOS neuromorphic chip with a scalable architecture for learning in networks of spiking neurons, in: *Custom Integrated Circuits Conference (CICC), 2011 IEEE*, Sept. 2011, 1-4
- Siekmann, Jörg: Künstliche Intelligenz: von den Anfängen in die Zukunft, in: Günther Cyranek / Wolfgang Coy (Hgg.): *Die maschinelle Kunst des Denkens, Perspektiven und Grenzen der künstlichen Intelligenz*, Braunschweig / Wiesbaden 1994, 11-41
- Sophokles: *Antigone*, in: Sophokles / Wilhelm Willige: *Dramen: griechisch und deutsch*. Hrsg. u. übers. von Wilhelm Willige, überarb. von Karl Bayer, mit Anmerk. u. e. Nachw. von Bernhard Zimmermann, München ³1995
- Steels, Luc: When are robots intelligent autonomous systems?, in: *Robotics and Autonomous Systems* 15 (1995), 3-9
- Stolleis, Michael: *Geschichte des öffentlichen Rechts in Deutschland*, München 1988, 90
- Sturma, Dieter: Über Personen, Künstliche Intelligenz und Robotik, in: Thomas Christaller (Hg.): *Autonome Maschinen*, Wiesbaden 2003, 38-55
- Sturma, Dieter: Was ist der Mensch? Kants vierte Frage und der Übergang von der philosophischen Anthropologie zur Philosophie der Person, In: Dietmar H. Heidemann / Kristina Engelhard (Hgg.): *Warum Kant heute?* Berlin 2004, 264-285
- Thrun, Sebastian: Toward a Framework for Human-Robot Interaction, in: *HUMAN COMPUTER INTERACTION* 19 (2004), 9-24
- Turing, Alan M.: Computing, Machinery and Intelligence, in: *Mind* 59/236 (1950), 433-46
- Volkelt, Johannes: *Kunstphilosophie und Metaphysik der Ästhetik*, München ²1924
- Weber, Max: *Wirtschaft und Gesellschaft*, hrsg. von Werner Gephart und Siegfried Hermes, Tübingen 2010
- Weiser, Mark: The Computer for the 21st Century, in: *Scientific American*, Sept. (1991), 94-104
- Weizenbaum, Joseph: Eliza - a computer program for the study of natural language communication between man and machine, in: *Communications of the ACM* 9/1 (1966), 36-45
- Weizenbaum, Joseph: *Die Macht der Computer und die Ohnmacht der Vernunft, aus dem amerikan. übers. von Udo Rennert (Computer Power and Human Reason. From Judgement to Calculation, 1976)*, Frankfurt a. M. 1977

Zwierlein, Eduard: Künstliche Intelligenz und Philosophie. Zur Debatte um J. R. Searle's Einwände gegen harte KI-Versionen, in: *Journal for General Philosophy of Science*, 21 (1990), 347-359