

Janina Loh

Roboterethik

Eine Einführung

Die Philosophin Janina Loh befasst sich in ihrem grundlegenden Buch mit den moralischen Herausforderungen, die beim Bau von Robotern und im Umgang mit ihnen eine Rolle spielen: Sind Roboter autonom? Können sie gar moralisch handeln? Haben sie einen moralischen Wert? Sollten ihnen Rechte zuerkannt werden? Wer ist zur Rechenschaft zu ziehen, wenn ein Roboter einen Menschen schädigt? Kritisch diskutiert Loh diese und weitere ethische Fragen und stellt die wichtigsten Lösungsansätze vor.

Janina Loh ist Universitätsassistentin im Bereich Technik- und Medienphilosophie an der Universität Wien.

Suhrkamp

*Meinen Gefährt*innen*

NORA TOBI
HANNAH WULF
FRANZI MIA
CHRISTIAN

Bibliografische Information der Deutschen Nationalbibliothek
Die Deutsche Nationalbibliothek verzeichnet diese Publikation
in der Deutschen Nationalbibliografie; detaillierte bibliografische Daten
sind im Internet über <http://dnb.dnb.de> abrufbar.

Erste Auflage 2019
suhrkamp taschenbuch wissenschaft 2277
© Suhrkamp Verlag Berlin 2019
Alle Rechte vorbehalten, insbesondere das
der Übersetzung, des öffentlichen Vortrags sowie
der Übertragung durch Rundfunk und Fernsehen,
auch einzelner Teile.
Kein Teil des Werkes darf in irgendeiner Form
(durch Fotografie, Mikrofilm oder andere Verfahren)
ohne schriftliche Genehmigung des Verlages reproduziert
oder unter Verwendung elektronischer Systeme
verarbeitet, vervielfältigt oder verbreitet werden.
Umschlag nach Entwürfen
von Willy Fleckhaus und Rolf Staudt
Druck: Druckhaus Nomos, Sinzheim
Printed in Germany
ISBN 978-3-518-29877-0

Inhalt

Einleitung	9
1. Bereiche der Robotik und ihre ethischen Fragen	19
2. Die Arbeitsfelder der Roboterethik	35
2.1 Roboter als moralische Handlungssubjekte	48
2.2 Roboter als moralische Handlungsobjekte	72
2.3 Inklusive Ansätze	95
2.4 Kritische Zwischenbilanz	121
3. Verantwortungszuschreibung in der Mensch-Roboter- Interaktion	126
3.1 Roboter als Verantwortungssubjekte	138
3.2 Roboter als Verantwortungsobjekte	162
3.3 Inklusive Ansätze der Verantwortungszuschreibung .	187
3.4 Kritische Zwischenbilanz	198
4. Abschließende Bemerkungen – Plädoyer für einen inkluisiven und kritischen Diskurs	204
Literaturverzeichnis	213

ANTHROPOMORPHISMUS	Turkle	Damiano und Dumouchel	Duffy
ANTHROPOMORPHISMUS	negativ	positiv	positiv
RECHTE FÜR ROBOTER	Nein	Ja	Ja
ZENTRISMUS	Anthropozentrismus	Anthropozentrismus, inklusiver Ansatz	Anthropozentrismus
Sind diese Roboter moralische Handlungsobjekte?			
KISMET	Nein	Nein	Nein
BLESSU-2	Nein	Nein	Nein
PARO	Nein	Ja	Ja
COG	Nein	Nein	Nein
iCUB	Nein	Ja	Ja

RECHTE FÜR ROBOTER	Darling	Gerdes	Tavani
ANTHROPOMORPHISMUS	positiv	positiv	positiv
RECHTE FÜR ROBOTER	Ja	Ja	Ja
ZENTRISMUS	Anthropozentrismus	Anthropozentrismus	Physiozentrismus
Sind diese Roboter moralische Handlungsobjekte?			
KISMET	Nein	Ja	Ja
BLESSU-2	Nein	Ja	Ja
PARO	Ja	Ja	Ja
COG	Nein	Ja	Ja
iCUB	Ja	Ja	Ja

ANSÄTZE MORAL PATIENCY	Sparrow	Levy	Johnson
ANTHROPOMORPHISMUS	neutral	neutral	neutral
RECHTE FÜR ROBOTER	ja (sofern <i>moral agents</i>)	ja (sofern <i>moral agents</i>)	neutral
ZENTRISMUS	Anthropozentrismus	Anthropozentrismus	Anthropozentrismus

Sind diese Roboter moralische Handlungsobjekte?			
KISMET	Nein	Nein	Ja
BLESSU-2	Nein	Nein	Ja
PARO	Nein	Nein	Ja
COG	Nein	Nein	Ja
iCUB	Nein	Nein	Ja

Tabelle 2: Positionen zur *moral patiency*

2.3 Inklusive Ansätze

Die innerhalb des dritten Bereichs der Roboterethik besprochenen Ansätze schlagen Alternativen vor zu dem aristotelischen Ethik-beziehungswise Bereichsethiken-Modell, dem damit einhergehenden (anthropologischen) Essenzialismus sowie dem häufig darin eingefassten (moralischen und epistemischen) Anthropozentrismus und stellen sich den philosophischen Herausforderungen zentristischer Ansätze im Allgemeinen. Dabei wird zum einen das tradierte Verständnis des moralischen Handlungssubjekts aufgebrochen, um es auch auf nichtmenschliche Wesen auszudehnen. Zum anderen sollen Kompetenzen, die im Rahmen der üblichen exklusiven Ansätze essenziellistisch einzelnen Akteurinnen und Akteuren zugeschrieben worden sind, nun relational als im Zusammenspiel unterschiedlicher menschlicher und nichtmenschlicher Wesen sich realisierend verstanden werden.

Wie einleitend in das Kapitel 2 bereits ausgeführt, ist das einende Vorhaben der auf den folgenden Seiten vorgestellten Perspektiven ein (moralisch und epistemisch) *inklusive* (also *inkludierendes*) Programm, das weder versucht, die Position der Menschen gegenüber anderen Wesen hervorhebend abzusichern, noch die Menschen mit anderen, ›niederen‹ Wesen gleichstellen will. Inklusive Ansätze wollen weder den Menschen moralisch relevante Eigenschaften absprechen noch ihre Rechte als Menschen beschränken. Sie vertreten den Anspruch, auf die teilweise umstrittenen Menschenbilder, die etwa den Menschenrechten und dem Humanismus zugrunde liegen, aufmerksam zu machen und über mögliche Alternativen nachzudenken.

Ebenfalls wurde bereits auf die Nähe inklusiver zu kritisch-posthumanistischen Positionen hingewiesen, was nun näher ausgeführt werden soll. Genauer lässt sich zunächst sagen, dass die kritisch-posthumanistischen Positionen eine Unterkategorie der inklusiven Theorien darstellen, insofern sie die Inklusion nichtmenschlicher Wesen in das moralische Universum vorrangig über eine Hinterfragung des humanistischen Paradigmas betreiben, das sie als grundlegend für das sogenannte westliche Menschen- und Weltbild sehen.

Dem kritischen Posthumanismus (Herbrechter 2018, 2009) ist nicht mehr primär an »dem« Menschen gelegen, sondern er hinterfragt die tradierten, zumeist humanistischen Dichotomien wie etwa Frau/Mann, Natur/Kultur oder Subjekt/Objekt, die zur Entstehung unseres gegenwärtigen Menschen- und Weltbilds maßgeblich beigetragen haben. Kritische Posthumanist*innen wie Donna Haraway, Karen Barad, Rosi Braidotti und Cary Wolfe möchten »den« Menschen überwinden, indem sie mit konventionellen Kategorien und dem mit ihnen einhergehenden Denken brechen, um so an einen philosophischen Standort hinter oder jenseits (»post«) eines spezifischen und für die Gegenwart essenziellen Verständnisses »des« Menschen zu gelangen. Aufgrund der fundamentalen Umwälzungen, die mit einer radikalen Hinterfragung des Humanismus einhergehen, werden in dieser Strömung darüber hinaus die üblichen Deutungen von menschlicher Gesellschaft und politischen Strukturen, von Wissenskulturen und deren Anspruch auf die Deutungshoheit von Tatsachen, Fakten und Wissen, des westlichen (und in der Regel weißen und männlichen) Kapitalismus und der Entwicklung der Massengesellschaft einer Totalrevision unterzogen (Braidotti 2016; Callus/Herbrechter 2013; Franklin 2009; Gane 2006; Krüger 2007; Nayar 2014). Fünf Kernelemente kennzeichnen das kritisch-posthumanistische Denken: (1) ein Ringen mit dem Humanismus,²² (2) eine Überwindung des Anthropozentrismus, (3) eine Hinterfragung des Essenzialismus und der (philosophischen) Anthropologie, (4) eine Kritik der Wissenskulturen sowie (5) ein deutlicher Appellcharakter sowie gesellschaftspolitische Implikationen (ausführlich in Loh 2018a).

Innerhalb des kritisch-posthumanistischen Paradigmas führt der

22 Für das dieser Studie zugrunde liegende Humanismusverständnis vgl. Loh 2018a: 17–31, 138–142.

argumentative Weg von einer traditionell essenzialistisch verfassten philosophischen Anthropologie über unterschiedliche Varianten kritisch-posthumanistischer Anthropologie in letzter Konsequenz schließlich zu einer radikalen »Alteritologie«, da in der kritisch-posthumanistischen Reflexion im eigentlichen Sinne nur Alteritäten existieren. Damit wäre der kritische Posthumanismus in seiner stringenten Ausformulierung ein umfassend inklusiver Ansatz. Denn er muss letztlich in einer Kritik nicht nur der humanistischen Anthropologie, sondern jeder anthropologischen Einhegung »des« Menschen münden und damit schließlich in einer Alteritologie, in der Vorstellung also, dass sich keine einzelnen Gattungen, Spezies und Rassen voneinander differenzieren lassen. Vielmehr, so die Theorie, leben wir in einem Universum, das von Alteritäten bevölkert ist, die so bunt und vielfältig sind, dass sie sich nicht in Schemata pressen lassen (Loh 2017c). Die gegenwärtigen kritisch-posthumanistischen Ansätze schlagen mehr oder minder moderate Realisierungen einer solchen radikalen Alteritologie vor.²³

Im Folgenden werden fünf inklusive Ansätze vorgestellt. Zwar legen sowohl Donna Haraway und Lucy Suchman den argumentativen Schwerpunkt gar nicht auf Roboter, sondern vielmehr auf alle möglichen Wesen. Dennoch lassen gerade sie sich als exemplarisch für diesen dritten roboterethischen Bereich verstehen, da sie mit der Neubestimmung des traditionellen moralischen Handlungsobjekts (vor allem in diesem Kapitel) sowie mit dem Abschied von einer essenzialistischen Konzentration auf die Relata einer jeweiligen Beziehung (insbesondere in Kapitel 3.3 am Beispiel der Verantwortung) ernst machen und dabei in besonderer Weise auch die Technik im Allgemeinen beziehungsweise moderne Technologien im Besonderen im Blick haben. Sie führen allgemein in das Thema der inklusiven Theorien ein, bevor dann mit Mark Coeckelbergh und David Gunkel zwei Vertreter explizit *roboterethischer* inklusiver Theorien vorgestellt werden. Sie konzentrieren

23 In der Einleitung wurde bereits darauf hingewiesen, dass den inklusiven Ansätzen nicht mehr vorrangig daran gelegen ist, die Speziesgrenzen durch die Definition der Essenz eines jeweiligen Wesens auszumachen. Relevanter als die Frage, ob (und warum) dieses Wesen etwa ein Mensch oder ein Roboter ist, ist es aus inklusiver Perspektive, eine gemeinsame Theorie und Praxis des guten Miteinanders zu finden. In meinem Habilitationsprojekt unternehme ich derzeit ebenfalls die Ausarbeitung eines inklusiven und kritisch-posthumanistischen Ansatzes.

sich direkt auf Roboter, nehmen allerdings keine weiteren Konkretisierungen vor, wohingegen es Luisa Damiano und Paul Dumouchel dann vorrangig um *soziale* Roboter geht. Damiano und Dumouchel verstehen zwar ihr Argument als Teil eines inklusiven Ansatzes, allerdings scheinen sie in mancherlei Hinsicht (vermutlich ungewollt) mit diesem zu brechen. Daher wird ihre Theorie zuletzt besprochen.

DONNA HARAWAY: Die promovierte Biologin und Naturwissenschaftshistorikerin verzahnt in ihren Schriften Philosophie, Feministische Theorie und Epistemologie, Literaturwissenschaft, Science Fiction, Technik- und Naturwissenschaften mit persönlichen Erfahrungen (Harrasser 2011). Auf diese Weise werden nicht nur Disziplingrenzen durchbrochen und unterschiedliche Wissensdiskurse miteinander verschmolzen, darüber hinaus erschwert diese dichte Collage den Zugang zu ihrem Werk. Bereits die erste ›inoffizielle‹ Schrift des kritischen Posthumanismus – Haraways *Cyborg Manifesto* (1985) – macht sich *zum einen* an eine Infragestellung der gegebenen Wissenskulturen und der akademisch-disziplinären Landschaft. In diesem Zusammenhang bereitet Haraway ein inklusives und relationales Verständnis von Verantwortung vor, das in Kapitel 3.3 diskutiert wird. *Zum anderen* (und darum geht es in diesem Kapitel primär) hinterfragt sie sowohl das gängige Subjektverständnis (und erweitert dieses auch auf nichtmenschliche Formen des Gegenübers) als auch den moralischen und epistemischen Anthropozentrismus. Die zahlreichen Figurationen, die ihr Werk insgesamt bevölkern und die sie als »family of displaced figures« (Penley/Ross 1991: 13) bezeichnet (etwa Cyborg, Modest Witness oder OncoMouse™; dazu Haraway 2000; Harrasser 2011: 586), stehen exemplarisch für diese Überwindung. Haraways Figurationen sind in kritisch-posthumanistischen Kreisen im Allgemeinen sehr beliebt, da sie in sich jeweils Dichotomien vereinen, um in der Anknüpfung an traditionelle Konzepte zugleich mit diesen zu brechen und auf diese Weise Wege zu einer Inklusion nichtmenschlicher Formen des Gegenübers in das moralische Universum zu eröffnen (Braidotti 2014: 167; Graham 2002; Harrasser 2011; Penley/Ross 1991: 20). Am Beispiel der Figuration der Cyborg wird nun exemplarisch aufgezeigt, wie Haraway vorgeht.

Im *Cyborg Manifesto* nennt Haraway Cyborgs »kybernetische Organismen, Hybride aus Maschine und Organismus, Geschöp-

fe der gesellschaftlichen Wirklichkeit und der Fiktion«. ²⁴ Cyborgs zeigen die Fragilität einer ernstzunehmenden Differenzierung zwischen Fakt und Fiktion auf, insofern »[g]esellschaftliche Wirklichkeit, d. h. gelebte soziale Beziehungen, unser wichtigstes politisches Konstrukt, eine weltverändernde Fiktion [ist]« (alle Zitate in 1995a [1985]: 33). Da sie sowohl imaginiert als auch real sind, verweisen Cyborgs darauf, dass die Grenze zwischen Fiktion und Realität konstruiert ist. Wir alle, so Haraway, haben diese Trennung zwischen Realität und Science Fiction, zwischen »Organismus und Maschine« in uns selbst aufgehoben, denn »wir sind Cyborgs« (beide Zitate in 1995a [1985]: 34). Auch die Unterscheidungen zwischen anderen wichtigen Dichotomien wie etwa Natur und Kultur, Tier und Mensch sowie »zwischen Physikalischem und Nichtphysikalischem« (1995a [1985]: 38) gehen uns verloren.

Diese und weitere Dichotomien, von denen die zwischen Realität und Fiktion sicherlich eine der grundlegendsten ist, stellen in Haraways Augen historisch gewachsene, von Menschen gemachte Narrative dar. Sie sind ohne Zweifel von enormer Relevanz für unser Selbstverständnis, unser Menschen- und Weltbild. Jedoch, so Haraway, bedeutet das nicht, dass es zu ihnen keine Alternative gäbe. Mit dem Anspruch, fundamentale Dichotomien aufzubrechen und in den Menschen als Cyborgs wieder zusammenzufügen beziehungsweise neu zu konstruieren, macht Haraway den Weg frei für weitere nichtmenschliche Akteurinnen und Akteure, die sich in allen ihren weiteren Werken finden. Seit den 1980er Jahren ist bei ihr eine Interessenverlagerung von techniknahen Themen und Figurationen hin zu tierischen Formen des Gegenübers zu beobachten. Insbesondere Affen beschäftigen sie in den 1990ern, Hunde und Tauben sind Gegenstand ihrer jüngsten Publikationen. Die Cyborgs sind darüber aber nicht vollständig vergessen, son-

24 Der von Manfred Clynes und Nathan Kline in »Cyborgs and Space« (1960) eingeführte Begriff sollte die Verschmelzung von einem menschlichen Organismus mit Technologien, die Menschen auf ein Leben im Weltraum vorbereiten, beschreiben. Ein Cyborg ist ein in spezifischer, d. h. in einer für das All und seine Herausforderungen modifizierter Weise veränderter Mensch. Das Kompositum Cyborg aus »cybernetic« und »organism« gibt deren organische Grundlage ebenso wenig eindeutig vor wie die Art und Weise ihrer technischen Modifizierung. Neben menschlichen Cyborgs sind tierische ebenso denkbar wie pflanzliche Cyborgs (Heilinger/Müller 2016).

dern werden von Haraway »mittlerweile als jüngere Geschwister in der viel größeren, queeren Familie der Gefährt*innenspezies« (2016 [2003]: 17) interpretiert, in der es »keine unabhängig existierenden Subjekte und Objekte und keine einzelnen Ursprünge, einheitlichen Akteur*innen oder abschließenden Ziele« (2016 [2003]: 13) gibt, sondern niemand der involvierten »Partner*innen vor dem Verhältnis oder der Bezugnahme existiert« (2016 [2003]: 18); »die Beziehung« [stellt] die kleinstmögliche Analyseeinheit [dar]« (2016 [2003]: 26, 31, 2008: 25-26).

In der Anerkennung unterdrückter und missachteter Formen des menschlichen und nichtmenschlichen Gegenübers werden die rassistisch-speziesistischen Wurzeln des *moralischen* Anthropozentrismus und Humanismus, leitbildgebend für insbesondere die sogenannten westlichen Kulturen mit ihrem vorrangig humanistischen, weißen, heterosexuellen und männlichen Menschenbild, freigelegt (Haraway 2008: 18). Der Versuch jedoch, einem solchen jeweiligen Gegenüber ein ›Gesicht‹, eine ›Stimme‹ zu geben, schlägt zuweilen in einen Paternalismus um – Haraway nennt diesen Vorgang eine »Semiotik der Repräsentation«. Wer nämlich etwa »für den Jaguar« oder »für den Fötus« zu sprechen sich berufen fühlt, der macht »das Objekt seiner Repräsentation [zur Verwirklichung der eigenen] kühnsten Träume«, der passiviert und instrumentalisiert das fragliche Gegenüber (alle Zitate in 1995b: 44-45). Das Paradigma des moralischen Anthropozentrismus lässt sich jedoch auch nicht einfach durch einen anderen Zentrismus ersetzen. So unterzieht Haraway in *When Species Meet* (2008) Jeremy Bentham's pathozentrischen Ansatz einer Kritik, wenn sie feststellt, dass die Frage nach der Leidensfähigkeit zwar unbenommen von großer Relevanz ist, ihr jedoch Überlegungen dazu, ob Tiere etwa spielen, arbeiten oder Freude empfinden können, in nichts nachstehen. Leidensfähigkeit, die Grundlage des Pathozentrismus, ist nicht die »ausschlaggebende Frage«, welche die »Ordnung der Dinge umkehrt« (beide Zitate in 2008: 22). An ihre Stelle können viele andere Attribute des menschlichen und tierischen Daseins treten, die für sich genommen jedoch wiederum auch nicht viel mehr als reduktionistische Engführungen darstellen – vorgenommen von den Menschen, die sich anmaßen, über Leiden, Freude, Spiel und Arbeit bei Tieren zu entscheiden.

Auch den *epistemischen* Anthropozentrismus nimmt sich Ha-

raway zu überwinden vor, was beispielsweise daran deutlich wird, dass sie ihre Hündin Cayenne ihre »Forschungspartnerin« (2018 [2016]: 16) nennt und Tauben als »kompetente Agenten [bezeichnet] [...], die einander und die Menschen zu situiert-sozialen, ökologischen, verhaltensbezogenen und kognitiven Praktiken befähigen« (2018 [2016]: 18).

Haraway formuliert ihren inklusiven Ansatz in *Unruhig bleiben* (2016) in Form einer Ethik der Verwandtschaft aus, die erst der Frage nach Verantwortung im eigentlichen Sinne gerecht werden könne (siehe Kapitel 3.3). Denn »[s]ich auf eigensinnige Art verwandt zu machen« schafft Bezüge und reicht weit über die engen Grenzen »der göttlichen, genealogischen und biogenetischen Familie« hinaus. Wer in eine jeweilige »Sippe« aufgenommen wird, mit wem wir uns jeweils verwandt machen, bleibt zwar offen und individuell festzulegen. Allerdings geht es Haraway an dieser Stelle vorrangig darum, die Grenzen dessen, was wir bislang als Familie oder Freundeskreis definiert haben (siehe auch die Überlegungen zur freundschaftlichen Beziehung in Kapitel 2.2), zu erweitern und nun auch offiziell jene Wesen ins moralische Universum aufzunehmen, die wir implizit schon zuweilen dort lokalisiert haben – wie etwa bestimmte Haus- und Nutztiere. Die »wilde Kategorie« der Verwandtschaft führt dazu, dass wir »unruhig [...] bleiben« (alle Zitate in 2018 [2016]: 10-11). Das bedeutet, dass wir die Widersprüchlichkeiten, die unaufhebbaren Spannungen unseres Daseins genießen und die Ironie darin sehen lernen, wie Haraway bereits im *Cyborg Manifesto* bemerkt: »Ironie handelt von Widersprüchen, die sich nicht – nicht einmal dialektisch – in ein größeres Ganzes auflösen lassen, und von der Spannung, unvereinbare Dinge beieinander zu halten, weil beide oder alle notwendig und wahr sind« (1995a [1985]: 33). Indem wir uns verwandt machen, können wir unruhig bleiben, also die notwendigen Spannungen, Dichotomien und Widersprüche anerkennen und lustvoll mit ihnen ringen. Denn nur weil sie von Menschen gemacht und gesellschaftlich konstruiert sind, heißt das noch lange nicht, dass Kategorien und Dichotomien nicht auch notwendig sind. Wir müssen im Alltag differenzieren, um handlungsfähig zu bleiben. Aber es ist eine offene Frage, *welche* Unterscheidungen wir *wann* treffen und *wen* wir dadurch aus dem Kreis der moralisch bedenkenswerten Wesen ausschließen. »Wir werden miteinander oder wir werden gar nicht«

(2018 [2016]: 11). Unruhig zu bleiben, »ist unsere Aufgabe« (2018 [2016]: 9), denn nur so kann auf Missstände hingewiesen, können Exklusionen erst zu einem kritischen Bewusstsein gebracht und als solche anerkannt werden.

LUCY SUCHMAN: Die Anthropologin und Wissenschaftlerin der Science and Technology Studies beschäftigt sich seit vielen Jahren mit der Mensch-Computer-Interaktion und insbesondere mit Militärtechnologien. In *Human-Machine Reconfigurations* (2007; der Neufassung ihres Buches *Plans and Situated Actions*) – entwickelt sie vor dem Hintergrund der Theorien von Donna Haraway, Bruno Latour, Judith Butler und anderen, aber vor allem mit Karen Barads kritisch-posthumanistischer »relationaler Ontologie« (Barad 2012: 18) einen Ansatz, um die Weisen, in denen wir uns zu anderen Menschen und Nichtmenschen (insbesondere Technologien) in Beziehung setzen und uns als autarke Wesen verstehen, zu *rekonfigurieren*. Mit ihrer relationalen Ontologie liefert Barad das kritisch-posthumanistische Fundament einer darauf zu gründenden Anthropologie und Ethik, für deren Ausformulierung Suchman im letzten Kapitel von *Human-Machine Reconfigurations* (2007) einen Vorschlag unterbreitet.

Barad schlägt mit ihrem agentiellen Realismus einen Weg zur Überwindung des Anthropozentrismus ein, indem sie »die dynamische Kraft der Materie« (2012: 40-41) anerkennt. Materie ist Barad zufolge nicht einfach nur passiv, sondern diskursiv, und der Diskurs ist nicht bloß aktiv, sondern auch materiell. Diese Sichtweise ermögliche es letztlich, nicht nur neue nichtmenschliche Subjekte als solche anzuerkennen, sondern auch die Dichotomie von Menschen und nichtmenschlichen Wesen radikal infrage zu stellen (2012: 30-32). In ihrem Gespräch mit Jennifer Sophia Theodor erläutert Barad das Vorgehen »eine[r] agentisch-realistische[n] Analyse« an den Beispielen der kalifornischen Rosine sowie des Fötus. Wenn wir eine kalifornische Rosine essen, so Barad, dann beißen wir zugleich in die »materiell-diskursive[n] Apparate« »wie Kapitalismus, Kolonialismus und Rassismus«, die an ihrer Produktion beteiligt waren (alle Zitate in 2015b: 184-186). Als Reaktion auf den Essentialismus formuliert sie einen relationalen Ansatz, der weniger die Relata, die Subjekte und Objekte, als vielmehr die Relation, also das, was zwischen ihnen liegt, in nichtessenzialistischer Weise in den Mittelpunkt stellt (siehe auch Coeckelbergh/Gunkel

2014: 722; Loh 2017c). Sie entwirft eine relationale Ontologie, in der keine singulären, autarken Akteur*innen existieren, sondern diese lediglich innerhalb ihrer gegenseitigen Verstrickungen anerkannt werden (Barad 2012: 18, 2007: 33, 93, 139).

Anders als Haraway, die insbesondere in ihren jüngeren Werken vorrangig Tiere behandelt, nimmt Suchman mit Blick auf Computertechnologien und Roboter das Vorhaben eines Neuverständnisses des »tradierten humanistischen« (2007: 270) Handlungssubjekts sowie der »Grenzen« zwischen Menschen und Maschinen in Angriff. Ähnlich wie bereits Haraway kommt sie dabei zu dem Schluss, dass nicht alle Differenzen per se hinfällig werden. Wohl aber können wir lernen, andere Unterscheidungen als die zu treffen, die uns das humanistische Menschenbild vorschlägt. Darüber hinaus, und darauf kommt es Suchman maßgeblich an, unterstützt uns das Hinterfragen des klassischen Verständnisses von Akteursschaft darin, die »Natur der Differenz« (2007: 260) selbst zu überdenken. Nicht nur, welche Unterschiede wir treffen, ist anthropologisch und ethisch von Belang, sondern wir müssen uns vergegenwärtigen und zur Diskussion stellen, wie wir Differenz selbst verstanden wissen wollen, was Differenz *ist*.

Entlang zahlreicher Beispiele aus den Technikwissenschaften, der Robotik und dem menschlichen Alltag, der durch Technologien fundamental geprägt und strukturiert ist, schlägt Suchman eine relationale Alternative zu der bislang essenzialistisch gedeuteten Mensch-Maschine-Beziehung vor, die auf klar definierten Grenzen und zuschreibbaren (und damit immer auch absprechbaren) Eigenschaften ruht. Ihr zufolge muss eine »Schnittstelle« nicht notwendig als »apriorische oder selbstverständliche Grenze zwischen Körpern und Maschinen gesetzt, sondern kann als Relation verstanden werden, die in bestimmten Situationen in Kraft ist und sich darüber hinaus mit der Zeit verschiebt« (2007: 263). Suchman würde Haraway darin zustimmen, dass es keine vorgegebene Weise der Rekonfiguration des Akteursverständnisses, der Dichotomien von Mensch und Maschine sowie der Kategorien, in denen wir unsere Beziehungen zu menschlichen und nichtmenschlichen Formen des Gegenübers definieren, geben kann. Die Art und Weise, in der wir Differenz verstehen und spezifische Unterscheidungen treffen, wirft »Fragen der Ausrichtung oder Verlagerung, Relation oder Entfremdung« auf, die unterschiedlich beantwortet werden

können. Diese Fragen sind der Mensch-Maschine-Interaktion nicht inhärent, »[s]ondern sie sind Effekte, die innerhalb von fassettenreichen Subjekt-Objekt-Sammlungen gelebt und erfahren werden« (beide Zitate in 2007: 266).

Wo nun Haraway ihr Konzept von Verwandtschaft als Grundlage einer relationalen Ethik des lustvollen Aushaltens von Spannungen in Anschlag bringt, gründet Suchman in baradscher Tradition ihre Ethik der Rekonfiguration auf der Annahme der »Intra-Aktion« (2011: 121) und verbindet damit auch die Neuverhandlung der Subjekt-Objekt-Dichotomie, wie sie in ihrem Text »Subject objects« (2011) an humanoiden Robotern exemplifiziert. Intra-Aktion setzt gerade nicht, wie die »Interaktion«, (mindestens) zwei autonome Entitäten voraus, die dann in eine Art Beziehung und »Austausch« miteinander treten (beide Zitate in 2007: 267). Intra-Aktion fußt vielmehr auf der Vorstellung von Interferenz (hier mit Bath u. a. 2013: 7 erklärt):

Interferenz ist der physikalische Begriff für die Überlagerung von Wellen. Er bezeichnet damit eine Form der Beziehung, die sich nicht übersetzen lässt in gängige Vorstellungen von Differenz und Identität, in denen abgeschlossene, stabile Entitäten miteinander in Beziehung treten. In der Interferenz wirkt die Welle, indem sie zu etwas anderem wird, zu einer anderen Welle, zu mehreren, zu kleineren, größeren Wellen, eventuell auch zu gar keinen Wellen. Wellen können sich nicht begegnen, ohne zu interferieren, ohne also zu etwas anderem zu werden. In diesem Sinne steht Interferenz für eine nichtbinäre, nichtlineare und dennoch relationale Form der Erzeugung von Andersheit, von Differenz, die in diesen Differenzen aber nicht aufgeht.

Aus dem Konzept der Intra-Aktion und damit implizit aus dem der Interferenz leitet Suchman die »wechselseitige Konstitution von Menschen und Artefakten« ab. Diese »Konstitution von Menschen und Artefakten tritt weder zu einer einzigen Zeit und an einem einzigen Ort auf, noch kreierte sie fixe Mensch-Artefakt-Relationen oder Entitäten« (beide Zitate in 2007: 268). Haraways Ethik der Verwandtschaft beruht demgegenüber noch eher auf dem Fundament von Interaktion und Identität, das Suchman mit der Vorstellung von einer wechselseitigen Konstitution in Interferenzbeziehungen überschreitet. So gelangt sie zu einer prozesshaften Idee von Menschen und Nichtmenschen, die sich in ihren ständigen Interferenzen dauernd miteinander und gegenseitig neu erschaffen

(2007: 278-279). »Handlungsfähigkeit [...] ist weder in uns noch in unseren Artefakten zu verorten, sondern in unseren Intra-Aktionen« (2007: 285).²⁵

Mit ihrer Ethik der wechselseitigen Konstitution von Menschen und Nichtmenschen intendiert Suchman (wie einleitend bereits bemerkt) durchaus keine *generelle* Aufhebung von Differenzen. Allerdings sollen wir überdenken, wie viel und was genau wir von unserem schweren humanistischen Erbe auch weiterhin als Grundlage unseres Menschen- und Weltbildes akzeptieren wollen. Wir benötigen, so Suchman, »ein Narrativ, das Menschen und Nichtmenschen zusammenbinden kann, ohne dabei die kulturell und historisch entstandenen Differenzen zwischen ihnen aufzuheben« (2007: 270). Es gilt, die »Zwillingsfallen eines kategorischen Essentialismus und einer Auslöschung von wichtigen Differenzen zu vermeiden« (2011: 137). Um auch nichtmenschliche Formen des Gegenübers in unser moralisches Universum inkludieren zu können, müssen wir allerdings in der Lage sein, ständig neu zu verhandelnde Formen der »Asymmetrie« oder besser noch »Dissymmetrie« (beide Zitate in 2007: 269) als Realisierungen der Interferenzen zwischen prozessual sich beständig neu entwerfenden Menschen und Nichtmenschen zu ertragen und zu unterstützen. Das wiederum klingt durchaus wie Haraways Aufforderung nach einem lustvollen »Aushalten von Spannungen« (Haraway 2018 [2016]). Suchman beschließt ihren Text »Subject objects« (2011: 138) daher mit der Aufforderung,

das wir uns von einer Tradition befreien, in der unser Interesse an Nichtmenschen entweder ihren reflektorischen oder kontrastierenden Eigenschaften gegenüber (einer bestimmten Figur von) unseren eigenen gilt, zugunsten einer Aufmerksamkeit für Ontologien, die radikal – aber immer kontingent – die Grenzen dessen, wo wir aufhören und der Rest beginnt, neu konfigurieren.

MARK COECKELBERGH: Bereits in seiner Habilitationsschrift *Growing Moral Relations* (2012) formuliert Coeckelbergh einen relationalen Ansatz mit Blick auf das Mensch-Tier-Verhältnis, den er zunächst mit dem Roboterethiker David Gunkel weiter ausarbei-

25 Ähnlich wie Haraway gelangt auch Suchman über dieses gewandelte Verständnis von Akteursschaft zu einem neuen Verständnis von Verantwortung, das in Kap. 3,3 näher in den Blick genommen wird.

ten wird, um ihn dann auf das Mensch-Maschine-Verhältnis anzuwenden. Coeckelbergh beginnt seine Untersuchungen mit einer Kritik der gewöhnlichen essentialistischen oder, wie er sie nennt, »Eigenschaftstheorien« (2014: 62) und setzt diesen, wie im Folgenden gezeigt wird, eine phänomenologische »relationale Ontologie« (2012: 45) entgegen, die eine »relationale Wende« (2010a: 209) vornimmt und die »moralische Signifikanz der Erscheinung« (2010b, 2009) auf der Grundlage »sozialer Relationen« (2010a: 219) begründet.

Die Standardposition der Eigenschaftstheorien wurde bereits in Kapitel 2.2 auf der Grundlage der Formulierung von Floridi und Sanders (2004: 350) vorgestellt und diskutiert: Jedes moralische Handlungssubjekt ist zugleich auch Objekt moralischen Handelns, indem es gerade der Status der *moral agency* gebietet, dieses Wesen auch in einer spezifischen (moralisch angemessenen) Weise zu behandeln. Sobald dem fraglichen Wesen jedoch die Bedingungen (in Form von Kompetenzen und Fähigkeiten) für *moral agency* abgesprochen werden, verliert es damit zugleich auch seinen Status als *moral patient*. Coeckelbergh geht es nun innerhalb des Horizonts der Standardposition insbesondere darum, dass der »moralische Status« einer Entität auf der Zuschreibung »einer spezifischen Eigenschaft« beruht. Bei der Standardposition handelt es sich aus seiner Sicht also insofern um eine essentialistische Theorie, als die Zugehörigkeit zu einer moralischen Gemeinschaft beziehungsweise der moralische Status eines Wesens an das Vorliegen bestimmter Attribute geknüpft ist (siehe auch Kapitel 2, Einleitung). In Coeckelberghs Worten: »Eine Entität x hat Eigenschaft p . Jede Entität, die Eigenschaft p hat, hat den moralischen Status s . Entität x hat den moralischen Status s .« Dort, wo es möglich ist, Eigenschaften zuzuschreiben, ist es ebenso möglich, sie abzuerkennen, erinnert uns Coeckelbergh mit Blick auf Sklaven, Frauen und Tiere (alle Zitate in 2014: 63).

Neben ihren (aufgrund ihres Essentialismus) implizit inhärenten exkludierenden Tendenzen gehen mit der Standardposition Coeckelbergh zufolge zusätzlich drei Herausforderungen einher. Zunächst zwei »epistemologische«: Wie sollen wir wissen, welche Eigenschaft für die Zuschreibung eines moralischen Status überhaupt hinreichend ist? Und auf welche Weise sollen wir eine fragliche Eigenschaft einem Wesen eindeutig zuschreiben? Überdies

können wir drittens die psychologische Tatsache einer »Diskrepanz zwischen intellektueller Argumentation und eigenem Erleben, zwischen Denken und Handeln, zwischen Glauben und Gefühl« nicht einfach ausblenden (alle Zitate in 2014: 63). De facto bauen wir beispielsweise zu bestimmten Maschinen eine emotionale Bindung auf und behandeln sie anders als andere artifizielle Systeme (siehe etwa die Roboterrobbe *Paro* in den Kapiteln 1 und 2.2) – ebenso wie viele Menschen zu ihren Haustieren eine ganz andere Beziehung haben als zu anderen Tieren. Der Standardposition zufolge steht uns gegenüber diesen und anderen nichtmenschlichen Wesen nur die Möglichkeit offen, ihnen zuzugestehen, mit den für einen spezifischen moralischen Status nötigen Kompetenzen und Fähigkeiten ausgestattet zu sein, oder aber den Menschen, die in ihnen moralische Subjekte erkennen, einen Fehler in der korrekten moralisch relevanten Einschätzung der fraglichen Entität vorzuwerfen.

Coeckelberghs Alternative zu der epistemologisch und psychologisch problematischen Standardposition liegt in einem relationalen Ansatz, der den moralischen Status als etwas begreift, was nicht essentialistisch einem Wesen zugeschrieben und damit immer auch wieder abgesprochen beziehungsweise von vornherein verweigert werden kann. Moralischer Status, so wie Coeckelbergh ihn versteht, entsteht in der Interaktion zwischen Wesen. Mit der Vorstellung, »dass eine Entität ohne Bezugnahme auf ihre Beziehungen nicht definiert werden kann«, steht er Haraways Ethik der Verwandtschaft näher als Suchmans Ansatz einer Intra-Aktion durch Interferenz, in dem die Subjekte selbst als prozessual sich zu dynamischen Einheiten zusammenfügen und wieder auflösen und ständig wandeln. Mit Haraway hält Coeckelbergh an der Vorstellung von klar definierbaren Relata, also handelnden Subjekten, fest, wenn diese auch nicht mehr dem humanistischen Ideal des autarken, unabhängig von seiner Umwelt existierenden (westlichen, weißen, männlichen) Menschen entsprechen. Wenn Coeckelbergh unter den Relationen insbesondere die »sozialen und natürlichen« benennt, wie etwa das »Ökosystem« eines Tieres und »die Netze sozialer Beziehungen« zu anderen Wesen (alle Zitate in 2014: 64), die in diese Umwelt eingebunden sind, klingt das hingegen nach genaueren Vorstellungen davon, mit welchen menschlichen und nichtmenschlichen Wesen wir Bindungen aufbauen (können), als bei Haraway, die dazu keinerlei Vorgaben macht (und vermutlich

beide Attribute des sozialen und natürlichen eher problematisieren würde). Haraway würde Coeckelbergh zwar vielleicht einschränkend dahingehend zustimmen, dass in unserem Kulturkreis die Anerkennung etwa eines Hundes als Gefährte wahrscheinlicher ist, als eine solch enge Beziehung zu beispielsweise einem Affen aufzubauen. Allerdings sind aus ihrer Sicht alle Relationen, die wir wählen, gleichermaßen natürlich, wichtig und einander moralisch ebenbürtig.²⁶

In seinem Text »Robot rights?« (2010a) schlägt Coeckelbergh mit Blick insbesondere auf die sozialen Beziehungen eine »relationale Wende« (2010a: 209) anhand dreier Kriterien vor: Erstens sei die »moralische Beachtung« eines Subjekts diesem nicht intrinsisch eingeschrieben, sondern extrinsisch in dessen sozialen Beziehungen und sozialem Kontext zu verorten. Zweitens seien diese moralisch relevanten Attribute als »sichtbar« zu verstehen, als »von-uns-wahrgenommene-Eigenschaften« (alle Zitate in 2010a: 214). Drittens schließlich sei die Wahrnehmung dieser Merkmale weder vom Kontext noch vom Subjekt unabhängig. Coeckelbergh schließt diesen Gedankengang in »Robot rights?« (2010a: 214) damit, dass

moralische Relevanz weder im Objekt noch im Subjekt liegt, sondern in der Relation zwischen beiden. Objekte wie beispielsweise Roboter existieren nicht allein im menschlichen Geist [...]; allerdings ist es auch wahr, dass wir nur dann Wissen von dem Objekt und seinen Merkmalen haben, wenn sie in unserem Bewusstsein auftauchen. Es gibt keinen direkten, unvermittelten Zugang zu dem Roboter als einer objektiven, Beobachter*in-unabhängigen Realität oder eines »Dings an sich«.

Mit Blick insbesondere auf den Roboter ist der soziale Kontext dessen Beziehungen zu anderen Wesen, seine »Situiertheit, Geschichte und Ort« und ebenso seine »natürliche, materielle, soziale und

²⁶ Jede exklusive (d. h. exkludierende) Position würde demgegenüber immer einige Beziehungen zu ausgewählten Wesen bevorzugen, etwa wenn die Beziehungen zu Menschen grundsätzlich als wichtiger angesehen werden als Beziehungen zu Tieren. *Diskriminierend* exklusive Ansätze wie ein Rassismus, Speziesismus oder Sexismus (Beispiele dafür wurden in Fußnote 3 der Einleitung in das zweite Kapitel) verfahren darüber hinaus offensiv und explizit benachteiligend und herabwürdigend gegenüber den als weniger wichtig eingestuft Beziehungen und Wesen. Inklusive Denker*innen sehen die Gefahr der Diskriminierung tendenziell in jeder exklusiven Theorie angelegt (vgl. Gunkels Einwand auf Wallachs und Allens funktionalistischen Ansatz in Kap. 2.1).

kulturelle Eingebettetheit«. Damit möchte Coeckelbergh keine neue Ontologie vorschlagen, die lediglich andere Eigenschaften als moralisch signifikant interpretiert, sondern im Gegenteil eine Ontologie, die die »menschliche Subjektivität« in seine relationale Ethik zu integrieren erlaubt. Der »moralische Status einer Entität ist konstruiert«, er ist von menschlicher Sprache, menschlichen Begriffen und menschlichem Denken abhängig, daher wandelbar und von Wesen zu Wesen unterschiedlich. Keine Zugangsweise und Beziehung zu einem spezifischen Roboter ist einer anderen *generell* vorzuziehen oder objektiv korrekt. Einige Relationen und Interpretationen sind, so Coeckelbergh, lediglich *situativ* »besser als andere« (alle Zitate in 2014: 64-65). Eine moralische Statuszuschreibung darf nicht im Vorhinein metaphysisch und essenzialistisch erfolgen.

In Übereinstimmung mit Haraway betont Coeckelbergh, dass sein relationaler Ansatz den Vorteil hat, »direkt die moralischen Fragen« zu adressieren, nämlich welche Beziehung wir zu einem anderen Wesen haben können, »wo wir [dabei] stehen und wer dieses »wir« ist« – und demzufolge, ob ein jeweiliges Gegenüber »Teil von »uns« ist« (alle Zitate in 2014: 74).

DAVID GUNKEL: Zuweilen in direkter Zusammenarbeit mit Coeckelbergh (etwa in dem gemeinsamen Text »Facing Animals«; 2014) entwickelt auch Gunkel einen inklusiven Ansatz; insbesondere in seinem Werk *The Machine Question* (2012). Gunkel stellt dort einleitend fest, dass wir René Descartes die Einschätzung von Tieren und Maschinen als geistlose Automaten, als von den Menschen radikal unterschiedene Alteritäten, zu verdanken haben (2012: 3). Ethisch schlägt sich diese Ontologie in einer Vorrangstellung »des« Menschen gegenüber allen möglichen Formen des nichtmenschlich »Anderen« und vor allem eben Tieren und Maschinen nieder, die diese Wesen aus dem moralischen Universum »ausschließt« (beide Zitate in 2007: 166). Erst im 20. Jahrhundert wurde diese ontologische (und folglich auch ethische) Unterordnung zumindest im Fall der Tiere zum Teil aufgehoben. Jetzt wäre es Gunkel zufolge höchste Zeit, auch mit Blick auf Maschinen dieselben Fragen zu stellen, wie wir sie uns hinsichtlich Tieren zu formulieren angewöhnt haben, insofern »[d]ie Frage nach der Maschine das Gegenstück zu der Frage nach dem Tier« (2012: 5) darstelle.

Als exemplarisch für die klassischen anthropozentrischen »Ethiken der Exklusion« (2007: 165) diskutiert Gunkel neben Descartes'

Ansatz auch die Positionen von Immanuel Kant, George Edward Moore und sogar Emmanuel Levinas, dessen »Ethik eine Anthropologie sowie eine spezifische Vorstellung von Humanismus voraussetzt« – so Gunkel (2007: 167). Diese und ähnliche traditionelle Theorien ließen sich zwar mitunter auf Tiere erweitern, allerdings ausschließlich unter Voraussetzung der anthropozentrischen und damit seiner Ansicht nach nichtmenschliche Formen des Gegenübers moralisch diskriminierenden Prämisse. Damit wendet sich Gunkel etwa gegen das Argument von Kant, der sich in § 17 des zweiten Teils der *Metaphysik der Sitten* (1797) zwar gegen Tierquälerei ausspricht, aber nur, da diese zu einer Verrohung, zu einer »Barbarisierung«, der Menschen führe, der einzigen Wesen, denen Kant zufolge ein Eigenwert zukommt, nicht aber, weil den Tieren ein Eigenwert zukäme. Tiere würden aus einer anthropozentrischen Perspektive also nur auf den ersten Blick moralisch ernst genommen.

Allerdings stellen alle anderen zentristischen Positionen (etwa der Pathozentrismus, der Biozentrismus und der Physiozentrismus; siehe Kapitel 2, Einleitung) Gunkel zufolge keine ernstzunehmenden Alternativen zu den exklusiven, also exkludierenden, anthropozentrischen Ethiken dar. Denn alle Zentrismen bleiben zum einen der essenzialistischen Standardposition verpflichtet und zum anderen dem der Subjekt-Objekt-Dichotomie impliziten Paternalismus, indem sie ein jeweiliges Erkenntnisobjekt definieren, das einem als Objekt gedachten Gegenüber Werte, Fähigkeiten und Kompetenzen zu- oder abspricht. Gunkel nennt diese anderen nichtanthropozentrischen zentristischen Ansätze »inklusive« (2007: 169) Theorien, da sie im Gegensatz zum Anthropozentrismus immerhin einigen nichtmenschlichen Wesen einen moralisch vergleichbaren Status geben.²⁷ In »Thinking otherwise« (2007) diskutiert er drei Positionen derartiger inklusiver Ethiken, die jedoch über den Essentialismus und Paternalismus der jedem zentristischen Denken implizit zugrunde liegenden Subjekt-Objekt-Dichotomie nicht hinauskämen: Michael Andersons, Susan Leigh Andersons und Chris Armens »Maschinenethik« (2007: 169), Ro-

²⁷ Nichtsdestotrotz nutzt er damit den Inklusionsbegriff anders als im Rahmen dieser Studie vorgeschlagen, wo alle zentristischen Ansätze als exklusive (denn auf die eine oder andere Weise exkludierende) Positionen bezeichnet werden (vgl. Kap. 2, Einleitung).

bert Sparrows »Androidenethik« und Luciano Floridis »Informationsethik« (beide Zitate in 2007: 171).

Obwohl im Rahmen der durch Anderson, Anderson und Armen vertretenen Maschinenethik der Kreis der moralisch bedenkenswerten Subjekte eine »signifikante Expansion« erfahre, sei sie »exklusiv an der Artikulation ethischer Richtlinien und Verfahren interessiert, die Maschinen den Umgang mit und ein Verhalten gegenüber Menschen vorgeben«. Es gehe dieser Position in letzter Konsequenz lediglich darum, »Menschen vor den potenziell gefährlichen Entscheidungen und Handlungen der Maschinen zu schützen« (alle Zitate in 2007: 170). Sparrow (siehe Kapitel 2.2) präsentiere mit seiner Androidenethik, die als kritische Weiterführung von Peter Singers Tierethik formuliert wird, eine »wichtige und fundamentale Herausforderung des anthropozentrischen Privilegs«. Jedoch »tut sie das, indem sie essenziell humanistische Werte nutzt und umgruppiert, die unglücklicherweise ihre eigenen Belange zu untergraben drohen« (beide Zitate in 2007: 171). Floridis Informationsethik schließlich, die sich auf *moral patients* und weniger auf *moral agents* konzentriert, geht Gunkel zufolge zwar weiter als jedes andere analytische Denken. Dennoch verbleibe auch sie immer noch innerhalb des Horizonts der »traditionellen binären Struktur« (2007: 172) der Subjekt-Objekt-Dichotomie zentristischer Ansätze. Selbst diese und ähnliche (aus seiner Sicht) inklusive (aus Sicht dieser Studie exklusive) Ethikvorhaben, wie sie etwa auch in der Umweltethik zu finden seien, bedürften in der Konsequenz immer noch der Vorstellung eines »Anderen« (2012: 160-161), den oder das sie moralisch diskriminieren, um auf diese Weise die eigene moralische Gemeinschaft zu schützen.

Gunkel nennt seine Alternative zu den herkömmlichen exklusiven anthropozentrischen sowie zu den jüngeren inklusiven, allerdings immer noch zentristischen, Positionen mit Rekurs auf dieses »Anderen«, das in das moralische Universum er sich aufzunehmen vornimmt, »anders denkend« (2012: 159, 2007). Er hat damit einen Ausbruch aus den bekannten Grenzen zentristischer Ansätze vor Augen und beruft sich dabei sowohl auf kontinentale Denker wie Jacques Derrida als auch auf analytische wie Thomas Kuhn (2012: 163). Gunkel zufolge besteht die Hauptschwierigkeit der genannten Theorien in der »Reduzierung von Differenz auf dasselbe« (2012: 162, 2007: 174), indem der Einschluss eines Gegenübers in

das moralische Universum durch die Leugnung oder Aufhebung dessen, was es anders macht, erfolgt.

Demgegenüber will Gunkel das Subjekt der herkömmlichen Ethiken dadurch »dezentralisieren« (2012: 163), dass es, wie er mit Allan Hansons Worten sagt, »mehr als Verb denn als Nomen wahrgenommen wird« (Hanson 2009: 98; siehe zu Hanson auch Kapitel 3.3). Damit steht Gunkel einerseits der Auffassung Suchmans eines prozessual sich je nach Kontext konfigurierenden Subjekts vermutlich näher als den Ansätzen von Haraway und Coeckelbergh, die immer noch zu einer eigenständigeren Vorstellung der miteinander in Beziehung tretenden Relata tendieren. Andererseits jedoch ähnelt das – mit der Annahme eines »sozial konstruierten, regulierten und zugeordneten moralischen Subjekts« (2012: 170) – durchaus Coeckelberghs relationaler Ethik sozialer Beziehungen. Und obwohl er Levinas' Theorie zuvor hinsichtlich deren vermuteter Anfälligkeit für einen humanistischen Anthropozentrismus kritisiert hatte, und obwohl er eingesteht, dass dieser zunächst den Eindruck vermittelt, ähnlich wie Floridi das moralische Handlungsobjekt in den Mittelpunkt seiner Überlegungen zu stellen, bereitet Levinas dennoch, so Gunkel, die Grundlage für eine »*Dekonstruktion* der konzeptuellen Ordnung von *agent* und *patient*« (2012: 176). Aus diesem Grund orientiert sich Gunkel in der Ausformulierung seines Ansatzes eines Andersdenkens letztlich explizit an Levinas. Sein Denken kreist dabei um drei Punkte:

Erstens müssen sich Philosoph*innen der Herausforderungen und sprachlichen Grenzen im Gebrauch der bekannten Termini bewusst sein (2012: 207-209). Weiterhin zwingt uns ein konsequentes Andersdenken Gunkel zufolge dazu, uns von der Vorstellung zu lösen, dass es eine definitive Antwort auf die »Frage nach der Maschine« geben kann. Vielmehr verweist uns die Maschine auf die Schwierigkeit des »Fundaments jeglicher Ethik« (beide Zitate in 2012: 211) und fordert einen beständigen Reflexionsvorgang im ethischen Urteilen darüber, über wen oder was in welcher Weise jeweils geurteilt wird. Schließlich dürfe diese »provisorische Natur der Ethik« nicht als »Versagen« derselben gewertet werden, dem ein wie auch immer zu definierender »Erfolg« gegenüberzustellen wäre (alle Zitate in 2012: 213). Ob und, falls ja, welche Maschinen in das moralische Universum zu integrieren sind, sei nicht ein für alle Mal festzulegen, sondern immer wieder neu zu verhandeln. So

stellt die Maschine die Weise, in der wir nach einem Gegenüber jeweils fragen, selbst durch ihre konkrete und situationsgebundene Existenz immer wieder in Frage.

LUISA DAMIANO UND PAUL DUMOUCHEL: Damiano und Dumouchel, deren Position bereits in Kapitel 2.2 mit Blick auf die Anthropomorphisierung (einiger spezifischer) artifizierender Systeme diskutiert wurde, entwerfen in ihrem Text »Anthropomorphism in Human-Robot Co-evolution« (2018) einen optimistischen Ansatz, der den Anthropomorphismus nicht als »kognitiven Fehler betrachtet«, sondern »als ein grundlegendes Werkzeug« (beide Zitate in 2018: 5), das in der Mensch-Roboter-Interaktion zum Einsatz gebracht werden und diese unterstützen und verbessern kann. Unter Berufung auf zahlreiche empirische Studien sehen Damiano und Dumouchel das Besondere sozialer Roboter darin, dass diese »die traditionellen ontologischen Kategorien tendenziell verwischen, die die Menschen für gewöhnlich zur Beschreibung der Welt nutzen«, allen voran die Subjekt-Objekt-Dichotomie, aber auch die Kategorien von belebt und unbelebt oder empfindungsfähig und nicht empfindungsfähig. Soziale Roboter sind ihnen zufolge immer »irgendwo dazwischen«, Roboter zu anthropomorphisieren ermögliche deshalb eine »neuartige Wissenschaft vom Menschen« (beide Zitate in 2018: 4). Den Ansatz der »synthetischen Ethik« (2018: 7), den sie vor diesem Hintergrund ausformulieren, hat zwei Ziele, nämlich (anthropologische) »Selbsterkenntnis« sowie »moralische Entwicklung« (beide Zitate in 2018: 8). Sie scheinen damit die Vorstellung von einer moralischen Entwicklung der Menschen durch die Anthropomorphisierung ihrer artifizierender Systeme zu vertreten.

Damiano und Dumouchel grenzen den Bereich der moralisch bedenkenswerten artifizierender Systeme auf den der sozialen Roboter ein. Ihren Überlegungen folgend, lässt sich so vergleichsweise leicht ausmachen, welche Roboter als *moral patients* zu identifizieren sind und welche nicht beziehungsweise mit welchen Maschinen etwa die Subjekt-Objekt-Differenzierung problematisiert wird. Allerdings vertreten sie den Anspruch, über die Aufhebung fundamentaler Dichotomien in der Anthropomorphisierung von (bestimmten) Robotern eine inklusive Theorie zu entwerfen. Dieser Anspruch soll auf den folgenden Seiten ernst genommen werden, wenn auch im Vergleich mit den zuvor in diesem Kapitel besprochenen Positionen bereits deutlich geworden sein sollte, dass sie durch die

essenzialistische Einschränkung ihrer Position auf soziale Roboter keinen strikten inklusiven Ansatz vertreten.

Damianos und Dumouchels Kritik der traditionellen Ethiken beläuft sich insbesondere auf deren negative Einschätzung, gar »radikale Verachtung aller sozialen Roboter und jeglicher Anthropomorphisierung von Technologien« (siehe zu dieser Haltung auch Kapitel 2.2). Hieraus folge in der Konsequenz eine »impotente Ethik«, die der faktischen Entwicklung sozialer Roboter nicht gerecht werden kann, da sie selbige verurteile und deren moralische Relevanz leugne. Demgegenüber sei ein »interaktionistischer verkörperter Ansatz« auszuformulieren, der jedoch die »irreduziblen (epistemologischen, phänomenologischen, operationalen et cetera) Differenzen, die Mensch-Roboter- von Mensch-Mensch-Interaktionen unterscheiden, anerkennt« (alle Zitate in 2018: 7). Spätestens damit zeigt sich, dass Damiano und Dumouchel ganz im Gegensatz zu genuin inklusiven Theorien immer noch an essenzialistisch zuschreibbaren Kriterien zur eindeutigen Definition der Spezies(grenzen) festhalten.

Die grundlegende »Differenz zwischen ›dichotomen‹ und ›synthetischen‹ Ethiken sei insbesondere in der Anerkennung der anthropologisch und ethisch relevanten Tatsache zu sehen, dass die Konstruktion sozialer Roboter ein Verstehen derselben und auch des menschlichen Wesens ermögliche. Mit der bewussten Anthropomorphisierung sozialer Roboter lernen die menschlichen Konstrukteurinnen und Konstrukteure, so Damiano und Dumouchel, nicht nur etwas über sich selbst, sondern mindestens ebenso viel über ihre artifiziellen Geschöpfe. So ermögliche beispielsweise die Einnahme einer synthetischen Perspektive die Einschätzung und entsprechende ethische Beurteilung von »Sexrobotern mit einer integrierten ›Vergewaltigungsoption‹ nicht als Simulation einer solchen, sondern als tatsächliche Vergewaltigung, da die »fragliche Praxis in einen sozialen Kontext eingebettet ist«. Durch ihren Verweis auf den sozialen Kontext, in den soziale Roboter mit ihren Interaktionen per se eingewoben seien, offenbart die Position von Damiano und Dumouchel eine gewisse Nähe zu Coeckelberghs Ansatz einer moralischen Statuszuschreibung in Abhängigkeit von natürlichen und sozialen Relationen. Gerade soziale Roboter gäben neue »Gesprächspartner*innen« ab, die mit der »Subjekt-Objekt-Unterscheidung brechen«, als »neue Kategorie von Interakteurin-

nen und -akteuren, mit denen neue soziale Beziehungen eingegangen werden können«. Am Beispiel der Sexroboter (siehe Kapitel 1) suchen Damiano und Dumouchel zu verdeutlichen, inwiefern die synthetische Perspektive im Bereich der sozialen Roboter die klare Differenzierung zwischen aktivem Subjekt einerseits und passivem Objekt andererseits aufhebt. Denn die Vergewaltigung eines Roboters sei immer noch Vergewaltigung, auch wenn die artifizielle Akteurin oder der artifizielle Akteur nicht in derselben Weise wie ein Mensch darauf reagiert« (alle Zitate in 2018: 8).

Für die ethische Bewertung der fraglichen Praxis spiele es letztlich keine Rolle – so die Konsequenz der synthetischen Ethik –, ob ein menschliches oder nichtmenschliches Gegenüber eindeutig als Subjekt oder als Objekt interpretiert werden kann. Die Analyse des jeweiligen Gegenübers in seinem spezifischen sozialen Kontext, unabhängig von dem es nicht gesehen und erst recht nicht verstanden werden kann, ist notwendig und hinreichend. Ganz ähnlich wie auch Gunkel gelangen Damiano und Dumouchel dabei zu dem Schluss, dass ihre synthetische Ethik »traditionelle Fragen, [die sich] im Fokus dichotomer Ansätze« befänden, nicht »ausschließen«, sondern »reformulieren« – mit Blick auf das emanzipatorische Potenzial sozialer Roboter, unsere Fähigkeit, Beziehungen einzugehen, zu »stärken« (alle Zitate in 2018: 8). Der Gegenstand der ethischen Beurteilung innerhalb des inklusiven Denkens verlagert sich von den Relata, die eindeutig als solche auszumachen sind, zu den Relationen, den eigentlichen Praktiken sowie dem sozialen Kontext.

BEISPIELE: Ebenso wie bereits in den Kapiteln 2.1 und 2.2 sollen nun die Beispiele aus der Robotik (*Kismet*, *BlessU-2*, *Paro*, *Cog* sowie *iCub*) besprochen und die fünf zuvor diskutierten Ansätze darauf übertragen werden, um etwaige Unterschiede zu den exklusiven Ansätzen einer Zuschreibung von *moral agency* und *moral patiency* auszumachen.

Mit ihrer Ethik der Verwandtschaft, die die Menschen lehrt, unruhig zu bleiben, de facto existierende Widersprüche und Spannungen also nicht nur auszuhalten, sondern lustvoll und ironisch mit ihnen zu ringen, erlaubt Haraway eine radikale Inklusion aller hier exemplarisch genannten Roboter in den Kreis der Gefährtinnen und Gefährten. Allerdings bleibt es von der jeweiligen individuellen Perspektive abhängig, ob man sich mit einem,

mehreren, allen oder keinem dieser artifiziellen Systeme vertrauen möchte. Etwa *Kismet* als einen Gefährten zu erkennen, hätte eine Aufhebung des epistemischen und moralischen Anthropozentrismus sowie in der Konsequenz jedes Zentrismus zur Folge. Von diesem Ergebnis lässt sich allerdings nicht ohne Weiteres darauf schließen, dass *Kismet* in jedem Kontext zum Beispiel einem menschlichen Kind moralisch gleichgestellt wäre und somit die Schädigung beider in derselben Weise zu beurteilen sei. Denn Haraway betont seit ihrem *Cyborg Manifesto* (1985) wiederholt, dass gesellschaftliche, politische, rechtliche und ökonomische Narrative nicht irrelevant sind, sondern unsere Welt und unser Zusammenleben de facto strukturieren. Wohl aber verpflichtet ihre Ethik der Verwandtschaft zu einem immer wieder geführten Diskurs über den gesellschaftlichen Common Sense einer Einschätzung und Bewertung beispielsweise konkreter artifizieller Systeme sowie der auf der Grundlage dieser geführten Diskurse formulierten Gesetze und Richtlinien. Ebenso wie es einmal dem gesellschaftlichen und wissenschaftlichen Standard entsprochen hat, anzunehmen, die Erde stünde im Mittelpunkt des Kosmos, könne auch die Interpretation der Menschen als einziger Wesen mit einem Eigenwert oder auch humanistische Dichotomien wie Natur/Kultur als Standardposition überdacht und gegebenenfalls verworfen werden. Eine solche (immer vorläufige) Einigung entbinde allerdings nicht das einzelne Individuum von seiner persönlichen Verpflichtung zu einer ethischen Positionierung. Solange sich Menschen innerhalb des Horizonts geltender Gesetze bewegen, erlaubt Haraways Ansatz bereits jetzt eine für jedes Individuum spezifische Definition des moralisch relevanten Gegenübers, was generell eine Inklusion in das moralische Universum und keine Exklusion zur Folge hat. Darüber hinaus lässt sich mit ihr der rechtliche und ethische Horizont einer Gesellschaft von innen heraus mit inklusiven (und nicht exklusiven) Intentionen in Frage stellen.

Suchmans Theorie einer über die Intra-Aktion (beziehungsweise Interferenz) wechselseitigen und beständigen Rekonfiguration der in eine Relation eingebundenen Wesen ermöglicht generell eine Inklusion der hier betrachteten Roboter in das moralische Universum. Allerdings sieht sie das gerade im Fall humanoider Roboter wie *Kismet* und *Cog* durchaus skeptisch. Denn hierüber würden essenziellistische anthropologische Vorstellungen sowie geschlech-

terbinäre Stereotype und die daran geknüpften Konzepte von Akteursschaft und Handlungsfähigkeit auf die fraglichen artifiziellen Systeme übertragen (2011: 122). Mit ihrer kritischen Einschätzung einer Anthropomorphisierung unterscheidet sich Suchmans Position deutlich von der Damianos und Dumouchels, die sich ja gerade einer generell positiven Erfassung des Potentials einer Anthropomorphisierung sozialer Roboter verschreiben. Suchman würde vermutlich darauf hinweisen, dass für eine Anerkennung dieser Roboter als echte nichtmenschliche Gegenüber eine Rekonfiguration der Mensch-Roboter-Interaktion sowie der beteiligten Relata nötig ist. Andernfalls stellten insbesondere die humanoiden Roboter essenziellistische und geschlechtsstereotypisierte Spiegelungen des westlichen und damit vorrangig weißen und männlichen Handlungssubjekts dar. Damit unterstreicht Suchman die persönliche Verantwortung in der Hinterfragung der eigenen Prinzipien, die in der oftmals impliziten Konzeption des moralischen (menschlichen und nichtmenschlichen) Gegenübers am Werk sind (siehe Kapitel 3.3).

Mit der Position, dass moralischer Status über die Interaktion, also über die sozialen und natürlichen Beziehungen, in die ein spezifisches Wesen jeweils eingebunden ist, zuschreibbar wird, lassen sich alle Beispielroboter, so ist mit Coeckelbergh anzunehmen, in das moralische Universum inkludieren. Abhängig sind die fraglichen artifiziellen Systeme dabei von ihrem jeweiligen sozialen Kontext und den menschlichen und nichtmenschlichen Wesen, die ihr Bezugsnetz bilden. Wo genau ein soziales Netz und damit die Möglichkeit, in Relation mit einer Entität zu treten, endet, wird aus seinen Ausführungen nicht deutlich, und ebenso wenig erschließt sich, inwiefern unterschiedliche Relata eines sozialen Kontexts sich gegenseitig unterschiedliche moralische Werte zusprechen können beziehungsweise welche Herausforderungen sich aus dieser Situation gegebenenfalls ergeben. Beides scheint aber generell zumindest für die Möglichkeit, dass ein nichtmenschliches Gegenüber (insbesondere ein Roboter) einen den Menschen äquivalenten moralischen Status erhält, nicht weiter relevant zu sein. Vermutlich könnte Coeckelbergh mit Haraway sagen, dass sich zumindest so lange kein Problem aus den möglicherweise differierenden moralischen Bewertungen ergibt, solange sich die urteilenden Individuen innerhalb des Horizonts des geltenden Rechts aufhalten. Seine Theorie

legt allerdings die Notwendigkeit offen, den geltenden rechtlichen Rahmen sowie den gesellschaftlichen und wissenschaftlichen Common Sense in der moralischen Beurteilung zu diskutieren und gegebenenfalls zu modifizieren.

Gunkels relationale Ethik des Andersdenkens ermöglicht die Einbindung der Beispielroboter in das moralische Universum über die drei seine Position rahmenden Prinzipien: die situationsgebundene Diskussion der hierfür relevanten Termini (wie etwa »Mensch«, »Roboter«, »moralisches Handlungssubjekt«), die Anerkennung der Herausforderung, sich der Frage nach dem moralischen Status von Robotern immer wieder neu zu stellen, sowie die Bereitschaft, in dem vorläufigen Status, den eine Beantwortung dieser Fragen jeweils nur erlangen kann, kein Versagen der Ethik zu sehen, sondern im Gegenteil die Akzeptanz derselben als ein soziales Narrativ, das zwar an sich notwendig für ein gesellschaftliches Zusammenleben ist, allerdings in einer jeweils historisch situieren sich artikulierenden Fassung immer kontingent bleibt. Vor diesem Hintergrund ist es möglich, in allen Beispielrobotern ein tatsächliches nichtmenschliches Gegenüber zu erkennen, das denselben moralischen Status hat wie Menschen.

Damiano und Dumouchel würden als einzige Vertreter*innen der in diesem Kapitel vorgestellten inklusiven Ansätze eine Einschränkung der Anerkennung nichtmenschlicher Formen des Gegenübers auf *soziale* Roboter vornehmen. An die in der Zusammenfassung von Kapitel 2.2 begonnenen Überlegungen anknüpfend, könnten sie von den gewählten Beispielrobotern vermutlich lediglich *Paro* und *iCub* als moralisch bedenkenswert einstufen, da nur diese als soziale Roboter im engen Sinne zu verstehen sind. Allerdings werfen sie mit ihrer Position einer synthetischen Ethik, die die jeweiligen Praktiken innerhalb eines sozialen Kontexts gegenüber den darin involvierten Relata in den Vordergrund rückt, zwei Fragen auf: zum einen die Frage, welche Roboter soziale Roboter sind, und zum anderen die Frage, warum die potenziell moralisch relevanten Formen des nichtmenschlichen Gegenübers auf die Kategorie der sozialen Roboter zu beschränken sind – oder anders: warum die essentialistischen Speziesdefinitionen und Bestimmungen der Speziesgrenzen letztlich doch nicht (vollständig) verworfen werden können.

ZUSAMMENFASSUNG: Es fällt auf, dass alle inklusiven Ansätze die

(moralische) Urteilskraft und Verantwortung des (menschlichen) Individuums als notwendiges Komplement zu den historisch spezifischen rechtlichen Regelungswerken sowie den gesellschaftlichen und wissenschaftlichen Narrativen ernst nehmen. Die Betonung der und des Einzelnen ist jedoch als Gegengewicht und nicht als Ersatz für die kollektive Position beziehungsweise überindividuelle Instanzen zu verstehen. Exemplarisch für diese Haltung steht die kritisch-posthumanistische Theorie Haraways. Haraway führt an zahlreichen Stellen ihres Werks ihre Haltung gegenüber den Methoden der Wissenserzeugung in den (Natur-)Wissenschaften aus, die sie »als eine spezifische Form des Erzählens, als eine kulturelle Praktik der Erzeugung von Bedeutungen« interpretiert (Hammer/Stieß 1995: 17; siehe auch Haraway 1995c: 105). Sie begegnet dieser mit dem Konzept des situierten Wissens, das veranschaulicht, inwiefern Wissen niemals kontextunabhängig, geschichtslos oder in irgendeiner Form »wahrhaft objektiv« ist. Ein radikaler Objektivismus würde einem blinden Glauben an die Autorität des geltenden Rechts oder der (Natur-)Wissenschaften folgen, also dem, was oben mit »kollektive Position« und »überindividuellen Instanzen« zusammengefasst wurde. Das Individuum muss vor dem Hintergrund dieser Haltung nicht selbst denken, keine eigenständige moralische Überzeugung ausbilden. Aus der Abkehr von einem radikalen Objektivismus resultiert Haraway zufolge jedoch keineswegs ein radikaler Konstruktivismus, der ihr auf zynische Weise relativistisch erscheint und vor allem zu politischem Handeln unfähig wäre (Penley/Ross 1991: 4). Die Position eines radikalen Konstruktivismus käme einer Überbetonung des (moralischen) Individuums gleich, das sich von keiner Instanz mehr kritisieren lassen muss (siehe hierzu auch Kapitel 3.3).

Alle hier diskutierten inklusiven Ansätze scheinen insofern eine ähnliche Auffassung von dem Verhältnis Individuum (radikaler Konstruktivismus) und Gesellschaft (radikaler Objektivismus) zu vertreten, als alle die moralische Funktion des Individuums als Gegenüber zu der nur scheinbar objektiven und zeitlosen Position gesellschaftlicher, wissenschaftlicher und rechtlicher Instanzen betonen. In allen Ansätzen vermisst man deshalb auffällig einen entsprechenden konkreten Vorschlag zur Ausbildung und Schulung der kritischen Urteilskraft und Fähigkeit zur Wahrnehmung der persönlichen Verantwortung in der moralischen Beurteilung eines

potenziellen menschlichen oder nichtmenschlichen Gegenübers als lokalisierbar im moralischen Universum. Was sich in den gängigen inklusiven Ansätzen mit Blick auf die Verantwortung in der Mensch-Roboter-Interaktion sagen lässt, wird Thema von Kapitel 3.3 sein.

Hier die Ergebnisse dieses Unterkapitels, zusammengefasst in Tabelle 3:

	Haraway	Suchman	Coeckelbergh	Gunkel	Damiano und Dumouchel
METHODE DER INKLUSION	unruhig bleiben	Intra-Aktion und Interferenz	soziale und natürliche Beziehungen	Diskussion von Termini, Wiederholung, Prima-facie-Ethik	moralische Beurteilung der sozialen Praktiken
ETHISCHER ANSATZ	Ethik der Verwandtschaft	Ethik der Rekonfiguration	Ethik der sozialen Interaktion	Ethik des Andersdenkens	Synthetische Ethik
EINSCHRÄNKUNGEN	Nein	hinsichtlich Anthropomorphisierung und (geschlechtlicher) Stereotypisierung	hinsichtlich eines jeweiligen sozialen Kontexts	Nein	ausschließlich soziale Roboter
Sind diese Roboter ein moralisches Gegenüber?					
KISMET	Ja	Ja, wenn auch zuvor Rekonfiguration	Ja	Ja	Nein
BLESSU-2					Nein
PARO					Ja
COG					Nein
iCUB					Ja

Tabelle 3: Inklusive Positionen

2.4 Kritische Zwischenbilanz

In Kapitel 2.1, in dem das Arbeitsfeld, das Roboter als potenzielle *moral agents* in den Blick nimmt, vorgestellt wurde, hat sich in der Diskussion der Ansätze von Floridi und Sanders, Sullins, Moor, Wallach und Allen sowie Misselhorn herausgestellt, dass

- *Autonomie* mindestens in einer negativen Definition generell eine Voraussetzung für die Zuschreibung von moralischer Akteursschaft darstellt.
- jene artifiziellen Systeme, die als potenzielle moralische Handlungssubjekte in Frage kommen, *bottom-up* (oder zumindest hybrid) konstruiert sind und damit in einem schwachen Sinn über *Lernfähigkeit* verfügen.
- fast immer Formen von *Intentionalität*, *Urteilkraft* und *Verantwortung* als zusätzliche Kompetenzen für die Annahme moralischer Akteursschaft eine Rolle spielen.
- die angestellten Überlegungen häufig auf einer implizit anerkannten *schwachen KI-These* fußen und damit die Simulation der fraglichen Voraussetzungen für (artifizielle) moralische Akteursschaft hinreichend zu sein scheint.
- alle Positionen implizit eine *anthropologische Prämisse* voraussetzen, indem sie von einem jeweiligen Menschenbild sowie einer Vorstellung des moralischen Handlungsobjekts ausgehend über die etwaige Anerkennung artifizieller moralischer Akteursschaft nachdenken.

In Kapitel 2.2, in dem ein Überblick über das Arbeitsfeld gegeben wurde, das sich mit Robotern als potenziellen *moral patients* befasst, konnte – anhand einer Besprechung der Ansätze von Turkle, Damiano und Dumouchel, Duffy, Darling, Gerdes, Tavani, Sparrow, Levy sowie Johnson – gezeigt werden, dass

- die Anerkennung eines Roboters als moralisches Handlungsobjekt häufig davon abhängig ist, dass dieser auch als moralisches Handlungsobjekt gesehen wird (die *Standardposition*).
- der häufig eingeschlagene Weg bei der Anerkennung eines Roboters als moralisches Handlungsobjekt über einen impliziten oder expliziten *Anthropozentrismus* verläuft.

- die Rechtfertigung einer emotionalen Bindung an eine Maschine sowie einer dadurch erfolgenden Wertzuschreibung und gleichzeitigen Inklusion derselben in das moralische Universum oftmals über den *Anthropomorphismus*, also die Vermenschlichung artifizieller Systeme, erfolgt.

Kapitel 2.3 schließlich, das eine Besprechung der inklusiven Ansätze von Haraway, Suchman, Coeckelbergh, Gunkel sowie Damiano und Dumouchel zu einem Neuverständnis des tradierten (moralischen) Handlungssubjekts sowie der diesem klassischerweise zugesprochenen Fähigkeiten und Kompetenzen umfasst, lässt sich darauf zuspitzen, dass

- kein neues Paradigma, keine neue Ontologie oder Sprache, die die alte ersetzen soll und damit alle Probleme derselben löst, gefunden werden kann, sondern dass inklusive Theorien oftmals als eine *Ergänzung zu den herkömmlichen Ansätzen* (in der Roboterethik) auftreten – wenn vor dem Hintergrund einer Akzeptanz des inklusiven Denkens auch radikale Änderungen unseres Menschen- und Weltbilds in Kauf zu nehmen sind.
- das *Zusammenspiel von individueller und kollektiver Perspektive* in der Einigung darauf, welche Formen des nichtmenschlichen Gegenübers (in diesem Fall insbesondere Roboter) in welcher Weise und unter welchen Umständen in das moralische Universum aufgenommen zu werden beanspruchen können, jeweils zu verhandeln und nicht als selbstverständlich vorauszusetzen ist.
- diejenigen, die sich im moralischen Universum aufhalten, dazu angehalten sind, eine eigenständige *moralische Position* zu beziehen, für die *Urteilkraft* benötigt wird und die die Wahrnehmung von persönlicher *Verantwortung* mit sich bringt.

Eine bislang nur flüchtig in Kapitel 2.2 in der Auseinandersetzung mit dem Anthropomorphismus angesprochene Schwierigkeit liegt in der Übertragung menschlicher Konzepte wie Autonomie, Lernfähigkeit oder Intentionalität auf Maschinen. Es bereitet häufig Unbehagen, nichtmenschliche Wesen mit Attributen zu belegen, die für gewöhnlich Menschen vorbehalten sind. Dieses Unbehagen ist gerechtfertigt, sofern die Gefahr besteht, dass damit der fraglichen Entität eine heimliche Teleologie, also Zweckgerichtetheit,

untergeschoben wird, wie das etwa in der Rede von konkurrierenden Genen der Fall ist (zahlreiche andere Beispiele aus den Naturwissenschaften ließen sich an dieser Stelle anführen). Gene konkurrieren nicht, Menschen konkurrieren. Gene haben keine Ziele, um derentwillen sie mit- und gegeneinander ringen, sondern Menschen haben Ziele, die sie in die Beobachtung von Naturprozessen in diese hineininterpretieren (siehe hierzu auch Immanuel Kant in seiner *Kritik der Urteilkraft*; 1790). Kritischen Posthumanist*innen zufolge geschieht das nicht nur in den Naturwissenschaften, sondern generell, da sich Subjekt und Objekt im Erkenntnisprozess nicht autark begegnen, sondern sich wechselseitig bedingen und quasi im Erkennen erst erschaffen. Allerdings macht es einen Unterschied, ob wir die Tatsache, dass wir notwendig immer durch unsere Augen zu sehen gezwungen sind,²⁸ als in den Kosmos gleichsam eingeschrieben sehen (das würde die Rede von konkurrierenden Genen rechtfertigen). Oder ob wir damit reflektiert und kritisch umgehen, immer bedacht auf die Konnotationen und Geschichten, die in begrifflichen Konzepten mitklingen, und stets in dem Bewusstsein, dass wir uns zu irgendeinem (historisch vermutlich nicht ausmachbaren) Zeitpunkt zu einem Gebrauch dieser Begriffe entschieden, auf eine für gewöhnlich in ihrer Nutzung antizipierte Definition geeinigt haben.

Unsere Sprache ist eine menschliche (so lange wir uns als Menschen verstehen), und es gibt keine Konzepte, sprachlichen Kategorien und Definitionen, die frei von jeglichen Konnotationen und unbeschwert von Geschichte und Tradition daherkommen. Mit Blick auf artifizielle Systeme etwa statt von Autonomie nun von Automation zu reden oder statt von Personen von Subjekten und Akteurinnen und Akteuren, macht die Sache nicht unbedingt besser oder einfacher. In manchen Fällen können wir von einem Begriffswechsel profitieren, wenn er uns eine neue Perspektive auf den Sachverhalt erlaubt, uns irritiert und die historische Last, mit der ein Konzept befrachtet ist, vor Augen führt.²⁹ Einen solchen

28 Was nicht mit dem epistemischen Anthropozentrismus gleichzusetzen ist, der die Annahme trifft, dass Menschen die einzigen zur Erkenntnis überhaupt befähigten Wesen darstellen.

29 Nebenbei ist das auch ein Argument für den Gebrauch des Gendersternchens: Es *soll* irritieren und uns auf die historische Last einer geschlechtlich eindeutig, nämlich männlich, kodierten Sprache aufmerksam machen!

Schritt versucht beispielsweise Sullins, wenn er in »When Is a Robot a Moral Agent?« (2006) den Begriff der Personalität in den Kontexten, in denen es um Roboter geht, vermeiden möchte. Es ist allerdings ein Fehlschluss, davon auszugehen, man könne mit der Neuwahl und, wie Suchman sagen würde, Rekonfiguration bestimmter Konzepte und sprachlicher Beziehungen eine neutrale Sprache entwickeln. Auf die Gefahren dieser Illusion weist Haraway in ihrem *Cyborg Manifesto* hin, wenn sie sagt, dass der »Traum einer gemeinsamen Sprache, wie alle Träume von einer perfekten, wahren Sprache, des perfekten getreuen Benennens der Erfahrung, ein totalisierender und imperialistischer Traum« ist (1995a [1985]: 51). Eine postulierte neutrale Sprache, eine unabhängige Beurteilungsinstanz, ist eine exklusive (und damit exkludierende) Autorität, die denselben argumentativen Stellenwert einnimmt wie Gott, »die« Geschichte, »die« Natur und »die« Wissenschaft. Der Traum einer gemeinsamen Sprache schließt alles aus, was, um es mit Gunkel auszudrücken, anders denkt, und kann dabei selbst doch nur dogmatisch sein, da es sich von außen (dessen Existenz per se geleugnet wird) nicht kritisieren lässt.

Sich von dem Traum einer objektiven Sprache zu verabschieden, bedeutet nicht zugleich die umfassende Leugnung jeglicher Form von Realität (siehe hierzu auch Fußnote 3 in der Einleitung zum zweiten Kapitel); der radikale Sozialkonstruktivismus ist nicht die einzige Alternative zu einem radikalen Objektivismus. Oder, wie Barad es mit Blick auf die Arbeit in den Naturwissenschaften zum Ausdruck bringt: »Zu sagen, dass etwas sozial konstruiert ist, bedeutet nicht, dass es nicht real ist« (2015a: 47). »Die Tatsache, dass naturwissenschaftliches Wissen sozial konstruiert ist, impliziert nicht, dass Naturwissenschaft nicht »funktioniert«, und die Tatsache, dass Naturwissenschaft »funktioniert«, bedeutet nicht, dass wir menschenunabhängige Fakten über die Natur entdeckt haben« (2015a: 10). Mit Blick auf artifizielle Systeme bedeutet das: Nur weil unsere Beschreibungsweisen von Robotern als etwa autonom, lernfähig oder mit Intentionalität ausgestattet keine »wahren« und objektiv (beziehungsweise empirisch) zu bestätigenden Aussagen über sie darstellen, ist es nicht vollkommen gleich oder egal, wie wir über Roboter sprechen und welche Kompetenzen wir ihnen etwa attestieren. Wir müssen uns der Verantwortung im Umgang mit den Begriffen, die wir nutzen, um Dinge zu beschreiben und

miteinander zu teilen, bewusst sein. Wenn wir also von Autonomie im Zusammenhang mit Robotern sprechen, sollten wir das reflektiert und kritisch tun und uns der Tatsache bewusst sein, dass die Genealogie der Autonomie unterschiedliche Definitionen erlaubt und die jeweils von uns befürwortete eine im Zweifel vorläufige ist, die weder historisch unabhängig noch »wahr« oder »richtig« ist oder den Anspruch erheben kann, von allen geteilt zu werden.

Die inklusiven (inkludierenden) Ansätze werfen den exklusiven (exkludierenden) im Allgemeinen vor, sie würden mit den ausnehmend problematischen, da implizit diskriminierenden, sexistischen, heteronormativen oder rassistischen Grundlagen ihres Denkens nicht transparent umgehen. Und da sie mit diesen ersten Prämissen ihres Denkens nicht transparent umgehen, werden diese dadurch implizit bestätigt und perpetuiert. Zuweilen würde zusätzlich sogar noch der Versuch unternommen, besagte Grundannahmen zu rechtfertigen, als objektiv und universell auszugeben. Die exklusiven Positionen hingegen werfen den inklusiven Denkerinnen und Denkern für gewöhnlich eine generelle Unschärfe vor, eine Verdunkelung der Umstände durch ein unnötiges und verwirrendes Aufweichen begrifflicher Kategorien und Speziesgrenzen. Die inklusiven Ansätze könnten höchstens zeigen, dass eine Ausformulierung jeglicher Ethiken der Verwandtschaft (Haraway), Ethiken der Rekonfiguration durch Intra-Aktion und Interferenz (Suchman), Ethiken der sozialen Interaktion (Coeckelbergh), Ethiken des Andersdenkens (Gunkel) sowie synthetische Ethiken (Damiano und Dumouchel) eine moralische Positionierung *erschwert*. Ihr Ruf nach einer Verantwortung des Individuums sowie einer Ausbildung seiner kritischen Urteilskraft könne den großen Herausforderungen, vor die uns die Einführung autonomer artifizieller Systeme in den unterschiedlichsten Kontexten und Nahbereichen des menschlichen Daseins stellt, indes nicht gerecht werden.

Im folgenden Kapitel wird zu zeigen sein, wie mit der Zuschreibung von Verantwortung in exklusiven und inklusiven roboterethischen Ansätzen jeweils umgegangen wird.