

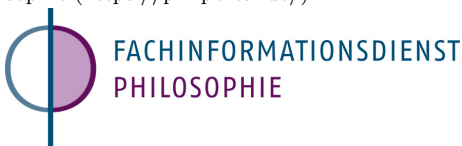
Philosophische Hefte

Peter Remmers

# Mensch-Roboter- Interaktion

λογος

Die Open-Access-Stellung der Datei erfolgte mit finanzieller Unterstützung des Fachinformationsdiensts Philosophie (<https://philportal.de/>)



Dieses Werk ist lizenziert unter der Creative Commons Attribution 4.0 Lizenz CC BY-SA (<https://creativecommons.org/licenses/by-sa/4.0/>). Die Bedingungen der Creative Commons-Lizenz gelten nur für Originalmaterial. Die Wiederverwendung von Material aus anderen Quellen (gekennzeichnet mit Quellenangabe) wie z.B. Schaubilder, Abbildungen, Fotos und Textauszüge erfordert ggf. weitere Nutzungsgenehmigungen durch den jeweiligen Rechteinhaber.



DOI: <https://doi.org/10.30819/4599>

# **Mensch-Roboter-Interaktion**

Peter Remmers

# Philosophische Hefte

*Band 3*

Herausgegeben von  
Prof. Dr. Thomas Gil

# Mensch-Roboter-Interaktion

Philosophische und ethische Perspektiven

Peter Remmers

Logos Verlag Berlin



## **Philosophische Hefte**

Herausgegeben von

**Prof. Dr. Thomas Gil**

Institut für Philosophie, Literatur-, Wissenschafts- und  
Technikgeschichte  
Technische Universität Berlin

Bibliografische Information der Deutschen  
Nationalbibliothek

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

© Copyright Logos Verlag Berlin GmbH 2018

Alle Rechte vorbehalten.

ISBN 978-3-8325-4599-4  
ISSN 2567-1758

Logos Verlag Berlin GmbH  
Comeniushof, Gubener Str. 47,  
10243 Berlin

Tel.: +49 (0)30 / 42 85 10 90

Fax: +49 (0)30 / 42 85 10 92

<http://www.logos-verlag.de>

# Inhaltsverzeichnis

Einleitung . . . . .	7
Was ist eine Mensch-Roboter-Interaktion?	13
Philosophische und ethische Aspekte der MRI . . . . .	15
Was ist ein Roboter? . . . . .	28
Konsequenzen für philosophische und ethische Aspekte . . . . .	49
Anmerkungen . . . . .	52
Literatur . . . . .	54





# Einleitung<sup>1</sup>

Die Zukunftstechnologie der Robotik entwickelt sich rasant. Sie lässt Visionen in greifbare Nähe rücken, in denen Roboter zu einer verbreiteten Arbeits- und Alltagstechnologie werden. Dabei kann man die gravierendste gegenwärtige Entwicklung der Robotik in einem Satz zusammenfassen: Roboter nehmen Kontakt auf. Hatten Roboter bisher ihren Platz hauptsächlich in unerreichbaren Science-Fiction-Szenarien oder hinter hohen Zäunen in der Industrie, so wird der praktische Einsatz von Robotern als Assistenten und Helferlein in unserer komplexen Lebenswelt immer öfter zur konkreten Realität. Roboter werden entwickelt und bereits eingesetzt u. a. für Assistenzfunktionen in Pflege und Rehabilitation, als autonome Fahrzeuge in Logistik und im Verkehr, als soziale Roboter im Dienstleistungsbereich und zur Unterhaltung sowie als Teleroboter in der Medizin, in Rettungseinsätzen und im Militär. In all diesen Anwendungsfeldern ergeben sich neuartige Schnittstellen zwischen Menschen und Robotern, die einige Fragen aufwerfen und weitreichende Diskussionen veranlassen.

Der Begriff „Mensch-Roboter-Interaktion“ (MRI; engl. Human-robot interaction, HRI) bezeichnet eine interdisziplinäre Forschungsrichtung, in der

es um das Verständnis, die Gestaltung und die Bewertung des Zusammenwirkens von Menschen und Robotern geht.<sup>2</sup> Beteiligte wissenschaftliche und technische Disziplinen sind u. a. Robotik, Psychologie, Soziologie oder Biologie. Ähnlich wie im allgemeineren Forschungsbereich der Mensch-Maschine-Interaktion geht es im weitesten Sinne um den Kontaktbereich zwischen Menschen und Maschinen, hier speziell zwischen Menschen und Robotern. MRI-spezifische Szenarien werden allerdings erst seit den 1990er Jahren im breiten Maßstab erforscht, zumal Roboter bis dahin hauptsächlich auf den vollautomatisierten Einsatz in der industriellen Produktion beschränkt waren, wo aus Sicherheitsgründen keinerlei Interaktion vorgesehen ist (abgesehen von Programmierung und Wartung). Fortschritte in der Robotik ermöglichen aber inzwischen Szenarien, in denen ein direkter Kontaktbereich zwischen Menschen und Robotern gegeben ist. Das Forschungsinteresse an solchen MRI-Szenarien kann zunächst roboterzentriert auf die technische Realisierung sicherer und effektiver Interaktionsformen konzentriert sein – ein Ziel, das mit enormen technischen Herausforderungen verbunden ist. Doch die Forschungsrichtung der MRI beschränkt sich nicht auf rein technische Fragen der Robotik, sondern thematisiert in benutzerzentrierten Ansätzen besonders die menschliche Seite

im Umgang mit Robotern. Hier kommen insbesondere psychologische und soziologische Untersuchungen der MRI zum Zuge. Im Unterschied zur technologischen Entwicklung geht es hier in erster Linie um die Gestaltung des Roboters und der Interaktion, wie sie sich für den menschlichen Interaktionsteilnehmer darstellt:

As it happens, if a non-researcher interacts with a robot that he or she has never encountered before, then what matters is how the robot looks, what it does, and how it interacts and communicates with the person. The ‚user‘ in such a context will not care much about the cognitive architecture that has been implemented, or the programming language that has been used, or the details of the mechanical design.<sup>3</sup>

Typische allgemeine Fragen, die im Rahmen der MRI-Forschung untersucht werden, sind beispielsweise die folgenden:

1. Wie wird ein Roboter vom Nutzer wahrgenommen, angenommen und infolgedessen verwendet?
2. Wie ist ein Roboter technisch zu gestalten, um den Zweck der Interaktion mit Menschen optimal zu erfüllen?

3. Welche Unterschiede und Gemeinsamkeiten bestehen zwischen Mensch-Roboter-Interaktionen und Mensch-Mensch-Interaktionen?
4. Wie können Roboter in komplexen Umgebungen der Lebenswelt eingesetzt werden?

Ein anschauliches Beispiel für ein vorrangiges Thema der MRI-Forschung betrifft den Aspekt der Sicherheit. Dabei geht es zunächst um den weitgehenden Ausschluss von Verletzungsgefahren im Kontaktbereich zwischen Menschen und Robotern. Doch neben der technischen Realisierung objektiver Sicherheit bildet auch die subjektive Sicherheit, d. h. die vom menschlichen Interaktionsteilnehmer wahrgenommene Sicherheit eine entscheidende Voraussetzung für gelingende Interaktion. Denn ein Roboter mag noch so sicher sein, wenn er beim Anwender ein unsicheres Gefühl oder gar Furcht auslöst, wird es wohl kaum zu effektiven Interaktionen kommen. Diese beispielhaften Bemerkungen sollen einen ersten Eindruck der MRI-Forschung vermitteln. Man sieht bereits, dass zur Beantwortung der genannten Fragen und zur Lösung beispielsweise von Sicherheitsproblemen technische und nicht-technische Disziplinen zusammenarbeiten müssen. Somit handelt es sich bei der MRI-Forschung um eine inter- und transdisziplinäre Unternehmung.

## Ziel und Ansatz des Beitrags

Im vorliegenden Beitrag geht es darum, einen Überblick über philosophische und ethische Aspekte der Mensch-Roboter-Interaktion zu vermitteln. Das bedeutet, dass es nicht unmittelbar um technische Lösungen, Sicherheitsaspekte oder überhaupt um empirische Erkenntnisse zum Verhältnis zwischen Robotern und Menschen gehen wird. Vielmehr geht es um grundsätzliche Fragen, die sich aus der prinzipiellen *Idee* einer Interaktion zwischen Menschen und Robotern ergeben. Zunächst wird daher erläutert, welche spezifischen Fragen in Philosophie und Ethik an die technologischen Konzepte der MRI gerichtet werden. Auf diesem Überblick über das Forschungsfeld aufbauend wird dann die Idee der Mensch-Roboter-Interaktion aus philosophischer Perspektive näher bestimmt und geklärt.

Die folgenden Ausführungen beziehen sich zu diesem Zweck zunächst auf den weiten Bereich von Roboter-Phänomenen, der im weiten Feld der Forschung, Entwicklung und Anwendung von Robotern gegeben ist. Allerdings ist neben den realen technologischen Phänomenen auch eine Besonderheit der Robotik zu berücksichtigen: Roboter und ähnliche Maschinen waren nämlich schon weit vor den tatsächlichen Realisierungen eine kulturelle Größe, insbesondere in Science-Fiction und in My-

then. Diese Einflüsse auf die Vorstellungen davon, was ein Roboter ist und wie sich entsprechende Interaktionen gestalten, sind nicht völlig von den realen Entwicklungen der Robotik zu trennen.

Eine philosophische Bezugnahme auf die MRI besteht aber nicht nur aus einer plausiblen Bestimmung charakteristischer Merkmale von Robotern, wie wir sie in verschiedenen realen und fiktiven Phänomenen vorfinden. Vielmehr ist mit einer Idee im philosophischen Sinne auch ein Anspruch auf Orientierung verbunden, so dass eine konstruktive Auseinandersetzung aus nicht-technischer Perspektive ermöglicht wird. In diesem Sinne soll eine These zum Charakteristikum von Robotern formuliert und erläutert werden, die insbesondere im Zusammenhang mit MRI-Szenarien fruchtbar erscheint: Roboter sind Maschinen, deren Funktionsweisen in Anlehnung an Handlungen interpretiert werden. Was das genau bedeutet und welche Konsequenzen sich aus dieser Charakterisierung ergeben, soll im Verlauf der folgenden Ausführungen deutlich werden.

## Was ist eine Mensch-Roboter-Interaktion?

Zunächst stellt sich die Frage, was mit dem Begriff der Interaktion im Konzept der MRI verbunden wird. „Interaktion“ bezeichnet zunächst ganz allgemein ein wechselseitiges aufeinander Einwirken von Akteuren oder Systemen. Im Verhältnis zwischen Menschen und Maschinen stehen dabei zunächst Aspekte der Kommunikation und der Steuerung, d. h. der Schnittstellen im Vordergrund. Im spezifischen Fall der Interaktion zwischen Menschen und Robotern können darüber hinaus weitere Interaktionsformen unterschieden werden.<sup>4</sup> Zunächst kann man von einer *kooperativen* Interaktion sprechen. Eine Kooperation zeichnet sich dadurch aus, dass beide Interaktionsteilnehmer auf ein gemeinsames und übergeordnetes Ziel hinarbeiten. Menschen und Robotern kommen dabei allerdings verschiedene Rollen zu, sie übernehmen jeweils verschiedene Teilaufgaben in der gemeinsamen Aktion. Ein Beispiel für kooperative Interaktion ist die Steuerung eines Roboters: Mit den Operationen des Roboters werden Aufgaben erfüllt, die dem Menschen beispielsweise schwere körperliche Arbeit abnehmen, während der Mensch die Programmierung oder die Supervision der Vorgänge übernimmt.

Eine Besonderheit der MRI ist die Möglichkeit *kollaborativer* Interaktionen. „Kollaboration beschreibt eine Interaktion als direkte Zusammenarbeit von Mensch und Roboter.“<sup>5</sup> Auch hier gibt es ein geteiltes übergeordnetes Ziel, aber im Gegensatz zur Kooperation werden auch Unterziele gemeinsam verfolgt. Ein anschauliches Beispiel ist die Übergabe eines Objekts vom Roboter zum Menschen: Der Roboter muss seine Bewegung auf den Übergabevorgang und auf den empfangenden Menschen abstimmen, ebenso wie der Mensch beispielsweise seine Hand aufhalten und dem Roboter entgegenhalten muss. Beide Interaktionsteilnehmer tragen auf derselben Ebene zum Gesamtvorgang bei. Dadurch entstehen offensichtlich neuartige Anforderungen an die Koordination zwischen Mensch und Roboter. Kollaborative Interaktionsformen markieren daher einen grundsätzlichen Unterschied zwischen MRI und der Mensch-Maschine-Interaktion im Allgemeinen.

Die Kategorien der kooperativen und der kollaborativen Interaktionsformen ordnen sich problemlos unter den allgemeinen Begriff der Interaktion ein, haben ihren spezifischen Sinn aber in der MRI. In Bezug auf Maschinen, die keine Roboter sind, erscheint die Unterscheidung dieser Interaktionsformen dagegen nicht unbedingt sinnvoll. Denn während im Verhältnis von Mensch und Maschine



im Allgemeinen die Steuerung der Maschine durch den Menschen im Vordergrund steht, erscheinen einige Operationen bestimmter Roboter nicht als Steuerungen, sondern vielmehr als direktes Zusammenwirken. Die Begrifflichkeit von Kooperation und Kollaboration bringt diesen Unterschied auf den Punkt.

An der Interaktion der Kollaboration wird zugleich das Prinzip der spezifischen Aspekte der MRI deutlich. Es geht um eine neuartige Weise, sich als Mensch auf technische Artefakte zu beziehen – in einigen Publikationen geht man sogar so weit, von Mensch-Roboter-Teams zu sprechen. Wie wir im Folgenden noch sehen werden, sind mit dieser Verschiebung der Interaktionsformen weitreichende philosophische und ethische Konsequenzen verbunden.

## **Philosophische und ethische Aspekte der MRI**

Für einen allgemeinen Überblick über die Themen, die sich in philosophischen Untersuchungen der MRI ergeben, ist zunächst zwischen theoretischen Fragen und ethischen Problemen zu unterscheiden. Das theoretische Interesse an der Technolo-

gie der Robotik und an der MRI orientiert sich hauptsächlich an zwei Zielen:

1. Hinter vielen philosophischen Beiträgen zur Robotik steht eigentlich ein anthropologisches Interesse, es geht also eigentlich darum, etwas über den Menschen herauszufinden. Roboter dienen vor diesem Hintergrund als kontrastives Werkzeug zur Untersuchung von bestimmten menschlichen Eigenschaften. So geht es beispielsweise um Klärungen von Aspekten der Verantwortung, der Körperlichkeit, des Bewusstseins, bestimmter Gefühle oder der Sterblichkeit – im Kontrast zu Robotern, die als technische Artefakte in das Umfeld dieser Themen zu rücken scheinen. Vor dem Hintergrund dieses Interesses können Roboter in letzter Konsequenz als Spiegelbild des Menschen thematisiert werden. Beispielhafte Fragen mit diesem Interesse können lauten:
  - Was erfahren wir aus den Entwicklungen und Zielen der Robotik über uns selbst als Menschen?
  - Inwiefern sind menschliche Eigenschaften an die spezifischen Merkmale menschlicher Existenz gebunden und unter wel-

chen Bedingungen können sie auf technische Artefakte übertragen werden?

- Wie ist mit den vielfältigen Übertragungen von Begriffen aus den ‚zwischenmenschlichen‘ Sprachspielen auf technologische Zusammenhänge und Interaktionen umzugehen?

2. Dagegen kann aber auch der Roboter als technisches Artefakt im Zentrum des Interesses stehen. Hier geht es allgemein ausgedrückt um die Frage, ob und wie Operationalisierungen von komplexen, an menschliche Eigenschaften angelehnten Fähigkeiten in technischen Artefakten gelingen können. So sind z. B. Kriterien angemessener Übertragungen vom nicht-technischen in den technischen Bereich zu formulieren, während die Merkmale der daraus entstehenden Mensch-Maschine-Systeme neu zu beurteilen und zu bewerten sind. Entsprechende Fragen sind z. B.:

- Welchen ontologischen Status haben die kollaborierenden Interaktionen zwischen Menschen und Robotern als technischen Artefakten – insbesondere im Kontext sozialer Interaktionen?

- Können die Bedingungen sozialer Interaktionen technisch angemessen realisiert werden – und wenn ja, wie?
- Unter welchen Umständen wäre einem künstlichen Akteur ein Bewusstsein und Gefühle zuzuschreiben?
- Welche Rolle spielen humanoide Gestaltungen von Robotern für die MRI?

In aktuellen Diskussionen über die Robotik bilden die ethischen Implikationen ein vorrangiges Thema. So werden im Rahmen der inzwischen etablierten *Roboter-Ethik* ethische Aspekte von MRI-Szenarien untersucht. Dabei finden sich Anknüpfungspunkte und Überschneidungen u. a. zur rechtswissenschaftlichen, sozialwissenschaftlichen und psychologischen Forschung, beispielsweise wenn es um Verantwortungsfragen, um Veränderungen der Arbeitswelt und um die Akzeptanz neuer Technologien geht. Charakteristisch für ethische Auseinandersetzungen im engeren Sinne sind normative Fragen, die nicht alleine durch den Verweis auf empirische Erkenntnisse geklärt werden können. In diesem Sinne kann (technik-)ethische Forschung in der Reflexion von Werten bestehen, die von der MRI betroffen sein können. Das Hauptziel ethischer Überlegungen besteht darin, transparente Grundlagen für Beurteilungen technologischer Gestaltung

zu vermitteln; unter Umständen können auch Gestaltungsmöglichkeiten und Lösungswege für eine ethisch verantwortungsvolle Technologieentwicklung aufgezeigt werden. Zum Überblick befasst sich die Roboter-Ethik u. a. mit den folgenden Problemen.

## **Das Problem der Autonomie**

Autonome Roboter operieren selbständig, d. h. sie reagieren in vergleichsweise komplexer Weise auf ihre Umwelt. Hier entsteht die Frage, nach welchen ethischen Kriterien die Reaktionen autonomer Roboter gestaltet werden können und sollen. In diesem Zusammenhang wird üblicherweise das sogenannte ‚Trolley-Dilemma‘ ins Spiel gebracht, ein Gedankenexperiment zur Veranschaulichung einer bestimmten Klasse ethischer Dilemmata. Je nach Variante des Dilemmas geht es dabei um eine Abwägung zwischen zwei ethisch problematischen Alternativen: Beispielsweise könnte ein autonomes Fahrzeug in die Situation gelangen, dass entweder eine Personengruppe auf der Fahrbahn verletzt wird, oder bei schnellem Ausweichen die Personengruppe im Fahrzeug gefährdet wird. In der ethischen Forschung dient dieses Gedankenexperiment zur Klärung der ethischen Grundsätze, auf die sich entsprechende Entscheidungen beru-

fen können. In Anwendung auf autonome Roboter kommen zwei weitere Aspekte hinzu: Erstens stellt sich die Frage, ob die Maßstäbe, an denen wir den Einsatz und die Funktionsweise technischer Artefakte ethisch beurteilen, zur Beurteilung autonomer Roboter angemessen sind. Beispielsweise wird in der Debatte um autonome Fahrzeuge häufig der Vergleich zum Verhalten menschlicher Fahrer gezogen; dabei ist allerdings zu beachten, dass an technische Artefakte grundsätzlich andere Anforderungen als an Personen gerichtet werden. In diesem Zusammenhang ist also der Übergang von verantwortungsvollem menschlichen Handeln zur Funktionsweise autonomer technischer Systeme zu reflektieren. Zweitens schwingt in den Debatten um eine ethische Dimension technischer Autonomie zugleich die Vorstellung mit, dass technische Lösungen zur generellen Verbesserung der Praxis auch in ethischer Hinsicht beitragen können. Beispielsweise könnte die Schnelligkeit technischer Datenverarbeitungen in kritischen Situationen angemessene Reaktionen ermöglichen, die das menschliche Reaktionsvermögen überfordern würden. So würden Entscheidungsspielräume eröffnet, die sonst aufgrund der menschlichen Natur schicksalhaft bleiben müssen. Charakteristische Fragen zu diesem Problemkomplex lauten:

- Wie sollen Roboter ‚entscheiden‘?

- Was erwarten wir in ethischer Hinsicht von autonomen Robotern, insbesondere im Vergleich zu unseren Ansprüchen an Menschen, die sich in vergleichbaren Situationen befinden?
- Wie ist in diesem Zusammenhang mit ethischen Dilemmata umzugehen?
- Können autonome Roboter unter bestimmten Umständen dazu beitragen, dass unsere ethischen Ideale besser erreichbar werden – und wenn ja, wie?

## **Das Problem der Verantwortung**

In engem Zusammenhang zum Problem der Autonomie stehen Fragen nach der Verantwortlichkeit für die Folgen, die durch die Interaktionen mit Robotern entstehen. Denn mit zunehmender Selbständigkeit von Robotern nimmt die Unklarheit über die Verantwortlichkeit zu. Während es bei offensichtlichen Fehlfunktionen noch möglich erscheint, die Verantwortung entweder den Entwicklern, den Herstellern oder den Anwendern zuzuordnen, kann der Spielraum der technischen Autonomie so komplex werden, dass Schäden nicht mehr vorhersehbar sind. Es kann folglich zu unerwünschten Aktionen eines Roboters kommen,

die nicht im Voraus als Fehlfunktionen oder Missbrauch deklariert wurden. Unter diesen Umständen kann möglicherweise niemand für die Aktionen autonomer Roboter verantwortlich gemacht werden. Darüber hinaus ist bei kollaborativen Interaktionen zwischen Menschen und Robotern mit einer Verteilung der Verantwortlichkeit zu rechnen. Wenn beispielsweise ein Roboter einem Menschen ein Objekt übergibt und das Objekt dabei beschädigt wird, dann ist der Mensch dafür sicherlich nicht in allen Fällen verantwortlich zu machen. Besonders dringlich werden diese Probleme im Paradigma des maschinellen Lernens: Da die Ergebnisse maschineller Lernprozesse nicht in jeder Hinsicht antizipiert und kontrolliert werden können, kann offenbar auch niemand mehr für die Folgen der Roboterfunktionen verantwortlich sein. Es stellen sich daher u. a. folgende Fragen:

- Wer ist verantwortlich für die Folgen, die aus dem Einsatz von autonomen Robotern resultieren?
- Wie verteilt sich Verantwortung zwischen ungleichen Interaktionspartnern in der MRI?
- Welche Konsequenzen ergeben sich aus dem Einsatz lernfähiger Roboter?
- Wie ist mit Leerstellen in der Verantwortungspraxis umzugehen?



## Das Problem der „moralischen Maschinen“

Bei diesem Problem handelt es sich gewissermaßen um eine Konsequenz aus den vorigen Problemen. Es geht um die Frage, ob und wie Roboter gleichsam als moralische Subjekte agieren können. Das Interesse an diesem Problem ist allerdings häufig eher theoretischer Natur, indem die allgemeinen Bedingungen für Moralität, ethisches Handeln und Verantwortung thematisiert werden. Eine künstliche Realisierung dieser Bedingungen und damit moralischer Subjektivität ist zwar mindestens auf dem aktuellen Stand der Technik kaum denkbar, rückt aber dennoch zunehmend in den Fokus der Aufmerksamkeit. In diesen Zusammenhang sind auch die vielzitierten Roboter-Gesetze einzuordnen, die zuerst im Werk Isaac Asimovs formuliert wurden:

1. Ein Roboter darf kein menschliches Wesen (wissentlich) verletzen oder durch Untätigkeit (wissentlich) zulassen, dass einem menschlichen Wesen Schaden zugefügt wird.
2. Ein Roboter muss den ihm von einem Menschen gegebenen Befehlen gehorchen – es sei denn, ein solcher Befehl würde mit Regel eins kollidieren.

3. Ein Roboter muss seine Existenz beschützen, solange dieser Schutz nicht mit Regel eins oder zwei kollidiert.

Abgesehen von der weitgehend offenen Interpretation dieser Gesetze ist offensichtlich, dass sie auf einem umfassenden inhaltlichen Vorverständnis aufbauen, dessen ‚Übersetzung‘ in technische Vorschriften unrealistisch ist. Der Adressat dieser Gesetze kann also zunächst nur der Mensch als moralisches Subjekt und als Entwickler von Robotern sein.<sup>6</sup> Im Kontext dieses Problemfeldes stellen sich beispielsweise folgende Fragen:

- Welche Voraussetzungen müssen erfüllt sein, damit Roboter nach moralischen Prinzipien agieren können?
- Welchen Prinzipien sollten sie folgen?
- Welche Konsequenzen hätte eine ‚moralische Maschine‘ für die menschliche Handlungspraxis?
- Ist es denkbar, unter bestimmten Umständen Roboter selbst für ihre Aktionen verantwortlich zu machen? Welcher Begriff von Verantwortung müsste dieser neuartigen Praxis zugrunde liegen?

## Das Problem des Sozialen in der Robotik

Wenn Roboter im sozialen Umfeld eingesetzt werden, dann kommen ihnen auch soziale Eigenschaften und Rollen zu. Das fängt bereits bei einer einfachen Namensgebung an: Ein freundlicher Name wirkt angenehmer als eine technische Typenbezeichnung. Diese und weitergehende soziale Funktionen können ausdrücklich erwünscht und gezielt eingesetzt werden. So gibt es neben Robotern als Spielzeuge für Kinder, künstlichen Haustieren und Sex-Robotern beispielsweise auch Assistenzroboter, die mit ‚sozialem‘ Profil ausgestattet werden, um etwa Interaktionen im privaten Raum möglichst einfach und angenehm zu gestalten. In Szenarien dieser Art findet eine vergleichsweise leicht zu realisierende Übertragung von menschlichen (oder auch menschlich-tierlichen) Eigenschaften und Kontexten auf technische Artefakte statt. Hier stellt sich die Frage, inwiefern diese Vermenschlichung der Technik problematisch werden kann und wo bestimmte Grenzen zu ziehen sind. Weitere Fragen lauten z. B.:

- Welche Rolle nehmen Roboter in sozialen Kontexten ein?
- Inwiefern sind soziale Funktionen von Robotern wünschenswert?

- Wie sind die Risiken zu bewerten, die durch affektive und emotionale Bindungen an Roboter entstehen?
- Ist das Vortäuschen von Sorge, Gefühlen u.ä. eines Roboters für soziale Zwecke problematisch?

## **Das Problem der Veränderung der Lebenswelt**

Auch wenn generell eine Anpassung neuer Technologien an die Bedürfnisse und Vorstellungen der Menschen angestrebt wird, hat dieser Prozess immer auch eine Kehrseite. Menschen werden sich immer auch in gewissem Grade auf die Technik einstellen müssen, sei es durch die Aneignung neuer Kenntnisse und Fähigkeiten im Umgang mit Technologien, sei es durch Umstellung der gewohnten Lebensweise bis hin zu Konflikten mit ethischer Tragweite. Eine Abschätzung der Technikfolgen aus ethischer Perspektive fragt nach Entwicklungen dieser Art und insbesondere nach den Werten, die davon betroffen sind. Grundsätzlich bringt der Mehrwert von Robotern immer auch Probleme in Form von neuen Risiken mit sich. Denn selbst, wenn beispielsweise größter Wert auf Sicherheit der MRI gelegt wird, sind neuartige Gefahren nicht völlig auszuschließen. Gerade die Entwicklung und

Anwendung von Robotern für ethisch besonders sensible Bereiche wie beispielsweise Pflege, Medizin oder Militär werden daher intensiv diskutiert. Je nach Art und Anwendungsbereich des Roboters sind verschiedene Werte betroffen, beispielsweise Selbstbestimmung, Unversehrtheit, Privatheit, Sicherheit oder Fürsorge.

Eng mit diesen Problemkomplexen hängen diejenigen Technikfolgen zusammen, die nicht nur mit unseren gegebenen Werten in Konflikt geraten können, sondern durch die sich Prinzipien unserer Bewertungen selbst anders gewichten oder sogar verändern. Wenn beispielsweise bestimmte Tätigkeiten nur noch von Robotern ausgeführt werden, könnten die damit verbundenen Fähigkeiten beim Menschen abnehmen und infolgedessen für unwichtig gehalten werden. Darüber hinaus könnte sich insbesondere in sozialen Interaktionen eine Art ‚Rückkoppelung‘ von Mensch-Roboter-Interaktionen auf den zwischenmenschlichen Umgang einstellen: Was im Umgang mit Robotern unproblematisch erscheint, im Umgang mit den Mitmenschen allerdings für Konflikte sorgen könnte, muss im gewohnten Verhalten auseinandergehalten werden. Es ergeben sich u. a. folgende Fragen:

- Welche Eingriffe in Lebenswelt und Gesellschaft sind durch den Einsatz von Robotern zu erwarten?

- Welche Veränderungen sind wünschenswert, welche nicht?
- Nach welchen Kriterien sind Wertekonflikte abzuwägen?
- Wie können Roboter und MRI-Szenarien unter dem Gesichtspunkt ethischer Werte gestaltet werden?

Schließlich sei noch darauf hingewiesen, dass neben den genannten philosophischen oder ethischen Fragestellungen, die sich aus dem spezifischen Themenkomplex der MRI ergeben, im Kontext der Robotik auch weitere ethische Probleme diskutiert werden, die nicht im engeren Sinne auf den Aspekt der Interaktion zwischen Menschen und Robotern abzielen. Hierzu zählen beispielsweise Fragen nach der gerechten Verteilung im Prozess der Technisierung der Lebenswelt, nach dem Wandel der Arbeit durch zunehmende Automatisierung sowie nach ethischen Prinzipien technologischer Forschung. Diese durchaus drängenden Themenkomplexe werden aus der vorliegenden Untersuchung ausgeblendet.

## **Was ist ein Roboter?**

Im Begriff des Roboters trifft aktuelle Technologie auf fiktionale, utopische und mythische Asso-

ziationen. Die daraus resultierende Komplexität dürfte wohl immer wieder zu phantastischen Erwartungen und Enttäuschungen führen. Entsprechend kompliziert gestaltet sich die Suche nach einer angemessenen Bestimmung dessen, was einen Roboter im Wesentlichen ausmacht. Dabei ist die Begriffsgeschichte des Roboters noch vergleichsweise jung: Das Wort ‚Roboter‘ wurde im Sinne von ‚Maschinenarbeiter‘ von dem tschechischen Schriftsteller Karel Čapek in seinem Drama *R.U.R. – Rossum’s Universal Robots* geprägt. Doch Idee und Mythos des Roboters als ‚Maschinen-Mensch‘ können zurückverfolgt werden bis in die Antike, wo beispielsweise von Talos erzählt wurde, einem gigantischen mechanischen Kämpfer, der von Hephaistos aus Bronze geschmiedet wurde.

Warum ist eine Begriffsbestimmung hier von Interesse? Für eine genauere Untersuchung der philosophischen und ethischen Implikationen der MRI benötigen wir zunächst ein klares Bild davon, was an den dargestellten Interaktionen besonders und neuartig ist. Und das Besondere und Neuartige der MRI ist sicherlich der Roboter, der erstmals überhaupt als realer Interaktionsteilnehmer auftritt. Daran anknüpfend kann man fragen: Was ist eigentlich ein Roboter? Gibt es ein spezifisches Merkmal, das Roboter im Unterschied zu anderen Maschinen besitzen? Und welche Besonderheiten ergeben sich

daraus für die Interaktion zwischen Menschen und Robotern?

Doch es geht im vorliegenden Beitrag nicht nur darum, den Gegenstand der MRI-Forschung näher zu bestimmen. Ein Ziel der philosophischen Arbeit soll auch darin bestehen, eine angemessene Perspektive auf das Thema aufzuzeigen, aus der heraus bestimmte Aspekte und Probleme erst deutlich werden. Es geht um Orientierung, also darum, die grundsätzliche Idee zu verstehen, die die Vorstellungen zur MRI anleitet. So soll schließlich eine Grundlage für einen reflektierten philosophischen und ethischen Zugang erarbeitet werden.

## **Eine technische Definition**

Zunächst erscheint es sinnvoll, eine bewährte technische Definition zurate zu ziehen:

Ein Roboter ist ein frei und wieder programmierbarer, multifunktionaler Manipulator mit mindestens drei unabhängigen Achsen, um Materialien, Teile, Werkzeuge oder spezielle Geräte auf programmierten, variablen Bahnen zu bewegen zur Erfüllung der verschiedensten Aufgaben.<sup>7</sup>



Auch wenn diese Definition schon 1990 formuliert wurde und insofern keine neueren Entwicklungen berücksichtigt, ist sie aus technologischer Perspektive immer noch akkurat. Sie kann als Grundlage für alle weiteren Spezifizierungen gelten, insbesondere sofern sie eine bestimmte technische Realisierung vorschreibt. Allerdings beschreibt sie in erster Linie Roboter, wie sie üblicherweise in der Industrie eingesetzt werden. Klassische Industrieroboter sind aber kein unmittelbarer Gegenstand der MRI-Forschung, denn gerade diejenigen Interaktionsformen, die für die MRI im engeren Sinne charakteristisch sind, kommen hier kaum vor. Vielmehr werden diese Roboter in räumlicher Distanz von Menschen und hinter Zäunen betrieben, um eine Gefährdung der menschlichen Mitarbeiter auszuschließen. Interaktionen in dieser Konstellation unterscheiden sich nicht grundsätzlich von Mensch-Maschine-Interaktionen im Allgemeinen – es geht dabei um Vorgänge wie Steuerung und Supervision. Zwar werden inzwischen auch vielfach kollaborierende Roboter für den Einsatz in der industriellen Produktion entwickelt und eingesetzt, aber zu den Besonderheiten dieser Roboter findet man in der VDI-Definition offensichtlich nichts. Als Ausgangspunkt für die MRI-Forschung und insbesondere für die philosophischen und ethischen Fragen an die MRI ist die Definition daher entsprechend zu ergänzen.

## Autonomie als definierendes Merkmal

Klassische Industrieroboter zeichnen sich durch komplexe Automation, aber nicht durch die Eigenschaft der Autonomie aus. Gerade diese Eigenschaft wird von einigen Autoren aber für wesentlich gehalten, wenn es um Roboter geht:

[W]e define ‚robot‘ as a *machine, situated in the world, that senses, thinks, and acts* [...]. [T]he generally accepted idea of a robot depends critically on the notion that it exhibits some degree of autonomy, or can ‚think‘ for itself, making its own decisions to act upon the environment.<sup>8</sup>

Mit dieser Formulierung schränkt George Bekey das Feld der Robotik auf die Aspekte der Autonomie und der Intelligenz (bzw. des ‚Denkens‘) ein. Nicht-autonome Maschinen, die z. B. ferngesteuert sein können, schließt er mit dieser Definition ausdrücklich aus dem begrifflichen Umfeld der Robotik aus: „[I]t does rule out as robots any *fully* remote-controlled machines, since those devices do not ‚think‘“.<sup>9</sup> Die Tatsache, dass Drohnen häufig als Roboter aufgefasst werden, erklärt Bekey am militärischen Beispiel folgendermaßen: „The U.S. Air Force’s Reaper unmanned aerial vehicle (UAV), though mostly teleoperated by humans, makes so-

me navigational decisions on its own and therefore would count as a robot.“<sup>10</sup>

Der Vorteil dieser Definition liegt darin, dass ein bestimmter ethischer Problemkomplex technologisch delegierter Entscheidungen deutlich wird: Wenn Roboter über Autonomie verfügen und Entscheidungen auf der Grundlage eines Verarbeitungsprozesses treffen, den wir mit menschlichen Denkprozessen vergleichen können, dann stellt sich hier offenbar die Frage nach den Bedingungen für ethisch ‚richtige‘ Entscheidungen. Damit wird eine Grundlage für die umfangreichen Debatten um das Problem der Autonomie gelegt, die sich um die Möglichkeiten ‚künstlicher moralischer Akteure‘ drehen. Roboter sind demnach ethisch relevant, wenn und weil sie mittels künstlicher Intelligenz selbständig Entscheidungen treffen. Fasst man Vorgänge dieser Art als technische Realisierungen menschlicher Grundbedingungen auf, dann fächern sich die oben dargestellten Problemfelder auf. Und auch aus technischer Perspektive ergeben sich einige Anschlussfragen: Wie können ethische Roboter programmiert werden? Können Kriterien ethischen Handelns im Paradigma des maschinellen Lernens erfüllt werden?

Diese Definition hat folglich den Vorteil, dass sie den Blick auf bestimmte philosophische und ethische Aspekte richtet. Allerdings hat sie auch offen-

sichtliche Nachteile, die sich an zwei Kritikpunkten festmachen lassen:

1. Der Begriff der Autonomie ist mehrdeutig und muss gerade in seiner Anwendung auf Roboter differenziert werden. *Technische Autonomie* als Eigenschaft von bestimmten Robotern „bezieht sich dabei auf die Eigenschaft einer Maschine, in bestimmten Bewegungsräumen Steuerungen und Aktionen auszuführen.“<sup>11</sup> Diese Art der Autonomie ist aber von der *personalen* Autonomie als Eigenschaft des Menschen zu unterscheiden: Sie „bezeichnet eine Fähigkeit von Personen, spontan Einstellungen einzunehmen und Handlungen auszuführen [...] vollzieht sich in der Form von Handlungen im Raum der Gründe.“ (Ibid.). Selbst, wenn eine Annäherung zwischen diesen beiden Autonomie-Auffassungen festgestellt oder hergestellt werden könnte, sollte diese Differenzierung einen Ausgangspunkt für weitere Überlegungen darstellen, insbesondere wenn es um ethische Probleme der Robotik geht. Als Bestandteil einer Definition erscheint der bloße Verweis auf die Eigenschaft der Autonomie ungenügend.
2. Die Eigenschaft der (technischen) Autonomie ist nicht charakteristisch für alle Gegenstände,

die wir unter dem Begriff des Roboters fassen. Es gibt einige Maschinen, die wir ganz selbstverständlich als Roboter bezeichnen, die aber nicht autonom agieren. Hier kommen beispielsweise die bereits angesprochenen klassischen Industrieroboter in Betracht, aber auch die weiten Felder der Telerobotik und der nicht-autonomen Assistenzroboter. Vor dem Hintergrund dieser unstrittig geläufigen Verwendung des Roboterbegriffs ist Bekeys Annahme, dass die allgemein anerkannte Idee des Roboters entscheidend von der Eigenschaft der Autonomie abhängig sei, unbefriedigend.<sup>12</sup>

## **Der Roboter als Werkzeug**

Ein reflektierter Ansatz zur Bestimmung des Roboterbegriffs findet sich in dem umfassenden Band von Christaller et al. (2001) zur Robotik:

Roboter sind sensumotorische Maschinen zur Erweiterung der menschlichen Handlungsfähigkeit. Sie bestehen aus mechatronischen Komponenten, Sensoren und rechnerbasierten Kontroll- und Steuerungsfunktionen. Die Komplexität eines Roboters unterscheidet sich deutlich von anderen Maschinen

durch die größere Anzahl von Freiheitsgraden und die Vielfalt und den Umfang seiner Verhaltensformen.<sup>13</sup>

In dieser Definition finden wir beide Tendenzen der oben diskutierten Definitionen wieder. Einerseits werden technische Eigenschaften genannt, die Roboter notwendig auszeichnen. Andererseits gibt die Definition eine Richtung für ein breiteres nicht-technisches Forschungsinteresse an: Die Betonung der „Erweiterung der menschlichen Handlungsfähigkeit“ und der Verweis auf „Verhaltensformen“ legt eine Anknüpfung an MRI-Forschungsfragen nahe. Mit dem Begriff der „Freiheitsgrade“ wird das von Bekey umrissene Problem der Autonomie außerdem terminologisch entschärft, indem der mehrdeutige Begriff der Autonomie differenziert und so die technologische Bedingtheit vermeintlicher ‚Entscheidungen‘ von Robotern verdeutlicht wird.

Der Band von Christaller et al. behandelt das Thema Robotik aus verschiedenen Disziplinen mit umfassendem Anspruch. Dieter Sturma formuliert dort in Kap. 4 eine grundsätzliche Kritik aus philosophischer Perspektive an einem vielfach zu beobachtenden naiven Umgang mit der Robotik: In der Wahrnehmung der Robotik wird der strikte Werkzeugcharakter des Roboters häufig durch anthropomorphistische Metaphern verwässert, so dass Vorstellungen und Erwartungen ins Spiel kommen, die

den technologischen Zielsetzungen und Entwicklungen nicht gerecht werden. Die Begrifflichkeit des Geistigen wird dann umstandslos auf technologische Realisierungen übertragen. Dies ist beispielsweise in Bekeys obiger Definition zu beobachten, wo der Begriff des Denkens (in Anführungszeichen) zur Beschreibung einer Eigenschaft von Robotern verwendet wird. Durch diese aus Sicht der Philosophie des Geistes und der Sprachphilosophie fragwürdigen Zuschreibungen werden Szenarien nahegelegt, in denen Roboter menschliche Tätigkeiten und Fähigkeiten ersetzen könnten. Gegen derartige Ersetzungsszenarien betont Sturma den strikten Werkzeugcharakter robotischer Funktionen, der keinen grenzüberschreitenden Einschnitt im Vergleich zu anderen technologischen Entwicklungen markiert.

Im Kontext von Sturmas kritischer Argumentation merkt man der erläuterten Definition an, dass sie anschlussfähig an philosophisch reflektierte Überlegungen ist. Sie umfasst einen weiten Bereich technologischer Phänomene und eröffnet interdisziplinäre Perspektiven für weitere Untersuchungen und kritische Einwände. Allerdings ist sie im Vergleich zu den beiden oben diskutierten Definitionen vergleichsweise unspezifisch: Als Beschreibung einer konkreten technischen Funktion ist die Angabe einer „Erweiterung der menschlichen Hand-

lungsfähigkeit“ zu weit gefasst, denn dieses Ziel wird mit nahezu allen technologischen Entwicklungen verfolgt.

Auch die Richtung, die die Definition und die kritischen Bemerkungen zur Terminologie für eine Anknüpfung philosophischer und ethischer Überlegungen einschlagen, ist in gewisser Hinsicht einseitig. Denn auch, wenn eine anthropomorphistische Begrifflichkeit und entsprechende Ersetzungsszenarien problematisch erscheinen, so ist diesem Problemkomplex wohl kaum mit einer stipulativen Definition beizukommen. Vielmehr zeigen u. a. psychologische und soziologische Forschungen zur MRI gerade die Schwierigkeit auf, Roboter ausschließlich hinsichtlich ihres Werkzeugcharakters zu betrachten. Spätestens bei der Erforschung *sozialer* Roboter muss dieser Anspruch über Bord geworfen werden. Doch es zeigen sich bereits bei der Gestaltung einiger kollaborativer Industrieroboter Tendenzen, die im Widerspruch zu einer bloßen Wahrnehmung als Werkzeug stehen und die sich wohl nicht durch eine nüchterne Betrachtungsweise eliminieren lassen.<sup>14</sup>

Fassen wir zusammen: Nicht-technische Auseinandersetzungen mit der Robotik konzentrieren sich häufig auf Eigenschaften wie Autonomie oder Intelligenz. Abgesehen von notwendigen Bedingungen wie der, dass die Operationen von Robotern



physisch realisiert sein müssen, sind die genannten Definitionen vor dem Hintergrund der vielfältigen Roboter-Phänomene allerdings häufig zu unspezifisch oder zu eng gefasst. Das gilt insbesondere vor dem Hintergrund des allgemeinen Sprachgebrauchs. Darüber hinaus benötigen wir für die Zwecke der vorliegenden Untersuchung eine Begriffsbestimmung, die uns eine umfassende und kritische *philosophische* Perspektive eröffnet. Wir suchen daher ein philosophisches Prinzip, eine Idee. Mit dieser Absicht möchte ich im Folgenden eine Charakterisierung des Roboterbegriffs vorschlagen, die Schwachstellen der genannten Definitionen beseitigt. Sie beansprucht Verfeinerungen der Begriffsbestimmung in zweierlei Hinsicht:

1. Die Vielfalt der verschiedenen Maschinen, die in Technologie und im Alltag mit dem Begriff des Roboters bezeichnet werden, wird einheitlich begründet.
2. Die philosophische Bedeutung der MRI soll klar und deutlich verständlich werden; es soll eine Stoßrichtung für weitere Überlegungen und Untersuchungen aufgezeigt werden.

## Roboter als handelnde Maschinen

Ein bestimmter Aspekt, an dem die philosophische Bedeutung der MRI verdeutlicht werden kann, wird in den diskutierten Definitionen des Roboters nicht aufgegriffen. So legt die kulturhistorische Herkunft von der Idee eines künstlichen Menschen es häufig nahe, sich Roboter in Form von humanoiden oder auch zoomorphen Gestaltungen vorzustellen. Diese Vorstellung ist in der heutigen technologischen Forschung nicht unbedingt leitend, da sich die Gestaltung häufig zunächst an den technischen Zwecken orientiert, so dass sich eine möglichst funktionale Morphologie als optimal herausstellt. Humanoide Formen sind für viele Zwecke dagegen sehr unpraktisch. Daher ist eine menschenähnliche Gestalt kein notwendiger Ausgangspunkt für die Charakterisierung von Robotern. Auch in der MRI ist die Interaktion zwischen Menschen und menschenähnlichen Robotern ein Spezialfall – denn eine funktionierende Interaktion muss nicht notwendig von gleichartigen Gestaltmerkmalen der Interaktionsteilnehmer abhängen.

Doch obwohl die äußere Gestaltung eines Roboters für die MRI nicht auf bestimmte Merkmale festgelegt sein muss, können wir ein ganz bestimmtes menschliches Merkmal sowohl in fiktionalen Darstellungen von Robotern als auch in realen Ent-

wicklungen der Robotik identifizieren: *Roboter sind Maschinen, deren Funktionsweisen in Anlehnung an Handlungen interpretiert werden.* Roboter operieren demnach in einem Kontext, der eine Auffassung ihrer Operationen als Handlungen zumindest begünstigt, im äußersten Falle sogar erfordert. Das gilt in besonderem Maße für MRI-Szenarien, kann aber in gewisser Hinsicht auf alle Arten von Robotern übertragen werden (siehe unten).

Was bedeutet das konkret? Die These lässt sich anhand einer naheliegenden kritischen Reaktion erläutern: Es scheint offensichtlich, dass Roboter als technische Artefakte selbstverständlich *nicht* handeln. Im Unterschied zu Menschen verfolgen Roboter streng genommen keine Absichten und sie haben keine übergeordneten Ziele. Diese Merkmale wären aus Sicht der gängigen handlungstheoretischen Positionen Voraussetzungen für ‚echte‘ Handlungen. Aus welchem Grund sollten also die Funktionsweisen von Robotern in Anlehnung an Handlungen interpretiert werden?<sup>15</sup>

Die Antwort lautet: Ob wir eine Bewegung als Handlung interpretieren, hängt in erster Linie vom Kontext ab, in den die Bewegung eingebettet ist. Im Unterschied zum handlungstheoretischen Verweis auf Absichten und Ziele gehen die praktischen Kriterien für Handlungszuschreibungen auf die Wahrnehmung konkreter Kontexte und passender Si-

tuationen zurück. Im Hinblick auf die Handlungstheorie sind Zuschreibungen und Erklärungen zu differenzieren: Während Absichten und Ziele eine Handlung erklären können, so stützt sich die Praxis der Handlungszuschreibung unmittelbar auf den öffentlichen und wahrnehmbaren Kontext, in dem die Handlung situiert wird. Der Verweis auf zugrundeliegende Absichten kommt gewissermaßen erst im Nachhinein auf der Reflexionsebene ins Spiel – der Kontext, der eine Interpretation einer Bewegung als Handlung ermöglicht, geht der Reflexion dagegen voraus.

Darin besteht das technologische Ziel und die Herausforderung der Entwicklung von Robotern für den Einsatz in MRI-Szenarien: Roboter müssen so gestaltet werden, dass sie möglichst unproblematisch in Handlungskontexte eingeordnet werden können. Dieses Ziel kann beispielsweise dadurch erreicht werden, dass der Roboter etwas tut, was an seiner Stelle normalerweise oder bisher ein Mensch getan hat; die geläufige Praxis legt es dann nahe, die Bewegungen des Roboters in Anlehnung an die ersetzte Handlung zu deuten. Aber selbst bei neuartigen Funktionsweisen, für die wir keinen Vergleichsmaßstab aus dem Bereich menschlicher Handlungen zurate ziehen, dürfte die besondere Komplexität der robotischen Operationen so zur Struktur

von Handlungen passen, dass eine Interpretation in Handlungskategorien passend erscheint.

Zum Verständnis dieses Vorgangs der Handlungsinterpretation kann auch eine Analogie zur Gestaltwahrnehmung dienen. Wenn wir beispielsweise zwei Punkte und einen Strich auf bestimmte Weise anordnen, kommen wir nicht umhin, diese Konstellation als Gesicht zu sehen. Aus dieser Wahrnehmungstendenz erklärt sich die Beliebtheit von Emoticons oder Emojis. Eine ganz ähnliche Tendenz, die allerdings nicht auf die rein visuelle Wahrnehmung beschränkt ist, bestimmt unsere Praxis der Zuschreibung von Handlungen: Es gehört zu unserer Auffassung eines Vorgangs als Handlung, dass er in einem handlungsrelevanten Kontext stattfindet, ein gewisses Bezugsobjekt hat und darüber hinaus möglicherweise weitere passende Merkmale aufweist (beispielsweise eine passende Bewegungsqualität). Sind diese Bedingungen erfüllt, dürften wir wie selbstverständlich geneigt sein, Handlungen zuzuschreiben.<sup>16</sup> Und gelingt es schließlich, diese Tendenz mit technischen Mitteln auszunutzen, dann kann uns die Handlungsqualität geradezu zwingend erscheinen. Eine humanoide Gestaltung eines Roboters *kann* diese Tendenz im Übrigen unterstützen, ist aber nicht das einzig denkbare Mittel.

Aus diesem Umstand erklärt sich schließlich auch die scheinbar metaphorische Begrifflichkeit, in der

wir über Roboter sprechen: Wir sagen, dass ein Roboter etwas Bestimmtes ‚tut‘; wir fragen nach ‚Entscheidungen‘ von Robotern; wir bezeichnen einen Roboter als ‚Akteur‘ und fragen in der Roboter-Ethik nach der Möglichkeit eines ‚moralischen Roboters‘. Wie selbstverständlich verwenden wir in MRI-Szenarien handlungsrelevante Begriffe zur Beschreibung der Bewegungen und Funktionsweisen eines Roboters. Diese Begriffsverwendungen sind zwar Übertragungen aus dem Bereich menschlicher Praxis, aber sie sind nicht beliebig, sondern werden durch die technologische Gestaltung suggeriert.

Die genaue philosophische Erklärung für diesen Interpretationsvorgang kann sich auf Suggestions-, Projektions- und Simulationskonzepte stützen.<sup>17</sup> Entscheidend ist an dieser Stelle vor allem, dass die Idee der MRI auf derartigen Interpretationspraktiken aufbaut und mitunter von gelingenden Handlungszuschreibungen abhängt. Betrachtet man beispielsweise die oben erläuterte Interaktionsform der Kollaboration, etwa wenn es um die Übergabe eines Objekts vom Roboter zum Menschen geht, so wird deutlich, dass der menschliche Interaktionsteilnehmer zu einem gewissen Grade handlungstypische Annahmen über den Roboter treffen muss. Es wäre wohl zu umständlich, eine langsam zunehmende Bewegung zu erwarten, in der ein Greifer um ca. 15-20 cm in meine Richtung gerückt wird – viel

einfacher ist es, einfach eine Übergabe des Objekts anzunehmen, und das bedeutet: eine Handlung zuzuschreiben.

Wir können also sagen, dass das technische Ziel eines Roboters in der MRI gerade darin besteht, Handlungen auszuführen; in diesem Sinne *sollen* die Operationen des Roboters als Handlungen interpretiert werden. Wenn das nicht gelingt, d. h. wenn wir als menschliche Interaktionsteilnehmer beim besten Willen keine Handlung erkennen können, dann wird auch die Interaktion scheitern. Mit anderen Worten: Die Idee der Handlung ist gewissermaßen in einen Roboter *eingebaut*. Daraus folgt insbesondere, dass der Handlungscharakter von Robotern nicht einfach als Täuschung oder als Fehlinterpretation abgelehnt werden kann. Der Unterschied zwischen ‚echten‘ Handlungen und ‚bloßen Interpretationen‘ von Bewegungen als Handlungen löst sich hier auf, weil die Deutungen nicht mehr nur subjektiv sind oder beliebig angepasst bzw. korrigiert werden können.

## **Gilt das Prinzip für *alle* Arten von Robotern?**

Es stellt sich allerdings die Frage, ob wir Interpretationen in Anlehnung an Handlungen auch bei verschiedenen Roboterarten erkennen können. Bei der

Diskussion der Definition, die sich auf die Eigenschaft der Autonomie beruft, haben wir bereits angemerkt, dass es offenbar Maschinen gibt, die wir Roboter nennen, obwohl keine nennenswerte Autonomie vorliegt. Aber können wir mit der Interpretation von Robotern als handelnden Maschinen alle Arten von Robotern erklären?

Zunächst kann die Bedeutung der *Autonomie* für die Robotik im Rahmen dieses Ansatzes erklärt werden. Denn Autonomie tritt beim Roboter als selbstständiges Operieren in Abhängigkeit von Informationen über die Umgebung auf. Mit diesem Merkmal ist aber offenbar schon eine große Ähnlichkeit zur Form von Handlungen verbunden. Das bedeutet: Die Deutung der Bewegungen eines Roboters als Handlungen wird durch die Eigenschaft der technischen Autonomie begünstigt.

Ähnlich deutlich liegt der Fall bei *sozialen* Robotern, mit denen soziale Interaktionen und Funktionen ermöglicht werden sollen. Kontext und Zweck dieser Anwendungen implizieren bereits eine Deutung der robotischen Funktionen als Handlungen, zumal hier nicht nur die Kommunikation mit anderen Menschen im Vordergrund steht, sondern der Roboter mehr oder weniger stark als soziales Gegenüber auftritt. Szenarien dieser Art lassen es geradezu unmöglich erscheinen, den Roboter nüchtern als rein technisch operierendes Werkzeug aufzufas-



sen; die soziale Struktur der Interaktion impliziert einen Handlungskontext.

Der Bereich der *Telerobotik* zeichnet sich durch die Fernsteuerung von Robotern aus. Zur Veranschaulichung können zwei Beispiele dienen: Zunächst gibt es Operationsroboter, die von Chirurgen ferngesteuert werden und deren Vorteil u. a. darin besteht, dass die zitternde Bewegung der Hände ‚herausgerechnet‘ wird. Bei dem zweiten Beispiel handelt es sich um Flugdrohnen, die auch im weitesten Sinne als Roboter aufgefasst werden. Zu beiden Beispielen können wir fragen: Warum nennen wir diese Vorrichtungen Roboter? Der Verweis auf teilautonome Funktionen liefert hier keine überzeugende Begründung, zumal wir manche technische Systeme mit vergleichbaren Funktionen nicht zu den Robotern zählen würden.

Nun hat die Interaktion der Fernsteuerung eines Roboters durch eine Person offenbar keine kollaborative, sondern eine kooperative Form. Trotzdem liegt es nahe, die Bewegungen einer ferngesteuerten Maschine als Handlungen zu interpretieren, da die Handlungen der steuernden Person hier räumlich *übertragen* werden. Drei Aspekte sind hier von Bedeutung: Erstens müssen die Operationen des ferngesteuerten Roboters so strukturiert sein, dass sie sich in den Rahmen eines Handlungskontextes einfügen. Wir müssten also geneigt sein, die

Bewegungen des Roboters als Handlungen zu interpretieren, auch wenn wir nichts von der Fernsteuerung wüssten. Zweitens realisiert der Roboter erst vollständig die Handlungen der fernsteuernden Person, d. h. der Roboter leistet einen entscheidenden Anteil zur Gesamthandlung. Zwar sind die Teilaufgaben voneinander unterschieden, aber es besteht ein Übertragungsverhältnis zwischen den Handlungskomponenten. Drittens besteht dieses Übertragungsverhältnis im Unterschied zur einfachen nicht-robotischen Fernbedienung darin, dass nicht nur technische Funktionen aus der Ferne ein- oder ausgeschaltet werden (wie es etwa bei der Fernbedienung eines Fernsehers der Fall ist), sondern dass bestimmte Bewegungen der Maschine direkt gesteuert werden. Daraus resultiert eine echte räumliche Übertragung einer Handlung, im Gegensatz zu einem Handeln, dessen Auswirkungen übertragen werden.

Schließlich erlaubt die dargestellte Analyse sogar eine Erklärung dafür, dass klassische *Industrieroboter* als Roboter gelten können. Zwar interpretieren wir die Bewegungen dieser Roboter nicht als Handlungen, dennoch können wir ihre Kategorisierung als Roboter damit begründen, dass sie aus dem paradigmatischen Handlungsorgan schlechthin bestehen: aus einem Arm. Der Begriff des Roboterarms impliziert eine Ähnlichkeit zum Handlungskontext,

wodurch die Bezeichnung als Roboter offenbar plausibel erscheint – ganz ohne Autonomie und im Gegensatz zu Waschmaschinen und Toastern.

## Konsequenzen für philosophische und ethische Aspekte

Folgen wir der dargestellten Analyse der Idee eines Roboters, dann erscheinen Roboter wie *Akteure* und insofern grundsätzlich menschenähnlich – in einigen Fällen auch unabhängig von ihrer äußerlichen Gestalt. Diese Auffassung steht im Widerspruch zu einem strikten Werkzeugcharakter, den man als aufgeklärter Philosoph für den vernünftigen Umgang mit Robotern einfordern möchte. Doch dieser innere Widerspruch zwischen Werkzeug und Handlungsinstanz ist der Idee des Roboters insbesondere in der MRI bereits eingeschrieben, er kann im Nachhinein nicht ohne Weiteres aufgelöst werden. Es ist dieser innere Zusammenhang, der die Idee des Roboters auch in ihren konkreten technischen Realisierungen philosophisch bedeutsam macht.

Doch der Verweis auf die Deutung der Operationen des Roboters als Handlungen dient nicht nur zum philosophischen Verständnis der verschiedenen Arten von Robotern, sondern auch als Perspektive auf die genannten ethischen Fragen der MRI. Denn

die dargestellte Auffassung liefert eine prinzipielle Begründung dafür, dass die Robotik als ethisch kritische Hochtechnologie angesehen wird. So kann beispielsweise die Rede von ‚Entscheidungen‘ autonomer Roboter verständlich gemacht werden: Es handelt sich dabei nicht um eine freie Metapher, sondern vielmehr um eine Übertragung, die durch Idee und Zweck des Roboters als handelnde Instanz nahegelegt wird.

Schließlich kann die Handlungsperspektive als Ausgangspunkt für eine kritische Positionierung zu den Entwicklungen in der Robotik dienen, indem darin die Unterschiede zwischen Robotern und Menschen klar werden. Die kritischen Ausgangsfragen lauten dann: Was folgt aus der Deutung technologischer Prozesse als Handlungen für unsere Vorstellung und unseren Begriff davon, was eine Handlung eigentlich ist? Welches Bild unserer Lebenspraxis und der von uns geschaffenen Technologien wird dadurch vermittelt? Ein anschaulicher Kontrast wird zum Beispiel am Problem der Verantwortung deutlich: Während menschliche Handlungen grundsätzlich mit Aspekten der Verantwortung verknüpft sind, werden diese Aspekte bei ‚künstlichen‘ Handlungen von Robotern gerade ausgegrenzt. Man kann es in aller Schärfe auch so ausdrücken: Das Problem der fehlenden Verantwortung bei autonomen Robotern entsteht deshalb, weil die Suggestion ei-

nes robotischen Handlungscharakters eben nur zu einem mangelhaften Ersatz führen kann.

Mit diesem Ansatz soll schließlich ein Schritt in die Richtung eines wesentlichen und wichtigen Ziels philosophischer und auch ethischer Forschung gegangen werden: nämlich eine klare und entschiedene Sicht auf die Verhältnisse zu erlangen, die wir durch neue Technologien erzeugen. Denn nur aus einer passenden Perspektive können wir diejenigen Aspekte deutlich erkennen, die uns bekennende Stellungnahmen und verantwortliche Gestaltungen der technischen Entwicklungen erlauben.

# Anmerkungen

<sup>1</sup>Der vorliegende Text wurde im Rahmen des BMBF-Begleitprojekts zur Ausschreibung *Autonome Roboter für Assistenzfunktionen – Interaktive Grundfertigkeiten (ARAIG)* verfasst. Für hilfreiche Anmerkungen und Hinweise danke ich Penelope Freund, Birgit Beck und Kurda Nejad.

<sup>2</sup>Für einen Überblick zur MRI-Forschung vgl. Goodrich & Schultz 2007 sowie Dautenhahn 2014.

<sup>3</sup>Dautenhahn 2014

<sup>4</sup>Vgl. Onnasch / Maier / Jürgensohn 2016.

<sup>5</sup>Onnasch / Maier / Jürgensohn 2016, S. 5

<sup>6</sup>Vgl. Anderson 2011.

<sup>7</sup>VDI-Richtlinie 2860, 1990

<sup>8</sup>Bekey 2012, S. 18

<sup>9</sup>Ibid.

<sup>10</sup>Ibid.

<sup>11</sup>Christaller et al. 2001, S. 126

<sup>12</sup>Der Verweis auf die autonomen Teilfunktionen von Drohnen zum Zweck der Navigation überzeugt nicht, zumal diese Eigenschaft wohl kaum die Begriffsverwendung anleiten dürfte. Unter dieser Voraussetzung wären alle Maschinen, die teilautonome Funktionen enthalten, zur Klasse der Roboter zu zählen, was sicherlich in einigen Fällen nicht einleuchten dürfte (vgl. z.B. den Fall einer Überwachungskamera, die bestimmte Bewegungen und Objekte fokussiert und verfolgt).

<sup>13</sup>Christaller et al. 2001, S. 19

<sup>14</sup>Vgl. z. B. Kuz et al. 2013.

<sup>15</sup>Vgl. Rammert / Schulz-Schaeffer 2002.

<sup>16</sup>Wie die Bedingungen für diese Tendenz konkret aussehen, ist letztendlich eine empirische Frage. Der philosophische Punkt besteht darin, dass grundsätzlich von externen und öffentlichen Kriterien der Handlungszuschreibung auszugehen ist.

<sup>17</sup>Vgl. z. B. Seibt 2017.

## Literatur

Anderson, Susan Leigh (2011): „The Unacceptability of Asimov’s Three Laws of Robotics as a Basis for Machine Ethics“, in: Michael Anderson / Susan Leigh Anderson (Hg.): *Machine Ethics*, Cambridge (Massachusetts): Cambridge University Press, S. 285-296.

Bekey, George A. (2012): „Current Trends in Robotics: Technology and Ethics“, in: Patrick Lin / Keith Abney / George A. Bekey (Hg.): *Robot Ethics. The Ethical and Social Implications of Robotics*, Cambridge (Massachusetts): MIT Press, S. 17-34.

Christaller, Thomas et al. (2001): *Robotik, Wissenschaftsethik und Technikfolgenbeurteilung* Bd. 14, Berlin / Heidelberg: Springer.

Dautenhahn, Kerstin (2014): „Human-Robot Interaction“, in: Mads Soegaard / Rikke Friis Dam (Hg.), *The Encyclopedia of Human-Computer Interaction*, Aarhus: The Interaction Design Foundation (2nd ed.) ([www.interaction-design.org/literature/book/the-encyclopedia-of-human-computer-interaction-2nd-ed/human-robot-interaction](http://www.interaction-design.org/literature/book/the-encyclopedia-of-human-computer-interaction-2nd-ed/human-robot-interaction), zuletzt abgerufen am 11.12.2017).

Goodrich, Michael A. / Schultz, Alan C. (2007): „Human-Robot Interaction: A Survey“, in: *Founda-*



*tions and Trends in Human-Computer Interaction* 1, Nr. 3, S. 203-275.

Kuz, Sinem / Mayer, Marcel Ph. / Müller, Simon / Schlick, Christopher M. (2013): „Using Anthropomorphism to Improve the Human-Machine Interaction in Industrial Environments (Part I)“, in: Vincent G. Duffy (Hg.): *Digital Human Modeling and Applications in Health, Safety, Ergonomics, and Risk Management. Human Body Modeling and Ergonomics* (DHM 2013), Berlin / Heidelberg: Springer, S. 76-85.

Onnasch, Linda / Maier, Xenia / Jürgensohn, Thomas (2016): „Mensch-Roboter-Interaktion – Eine Taxonomie für alle Anwendungsfälle“, in: *baua: Fokus*, S. 1-12 (DOI: 10.21934/baua:fokus20160630).

Rammert, Werner / Schulz-Schaeffer, Ingo (Hg.) (2002): *Können Maschinen handeln? Soziologische Beiträge zum Verhältnis von Mensch und Technik*, Frankfurt am Main: Campus.

Seibt, Johanna (2017): „Towards an Ontology of Simulated Social Interaction: Varieties of the ‚As If‘ for Robots and Humans“, in: Johanna Seibt / Raul Hakli (Hg.): *Sociality and Normativity for Robots. Philosophical Inquiries into Human-Robot Interactions*, Berlin/Heidelberg: Springer, S. 11-39.

Roboter sind Maschinen, deren Funktionen in Anlehnung an Handlungen interpretiert werden. Dadurch erscheint ihr praktischer Einsatz als Assistenten und Helferlein in unserer komplexen Lebenswelt möglich. Doch neben den technischen Herausforderungen ergeben sich neue philosophische und ethische Fragen.

**Logos Verlag Berlin**

**ISBN 978-3-8325-4599-4**

**ISSN 2567-1758**