



Ethische und soziologische Aspekte der Mensch-Roboter-Interaktion

baua: Bericht

**Forschung
Projekt F 2369**

L. Onnasch
T. Jürgensohn
P. Remmers
C. Asmuth

**Ethische und soziologische Aspekte der
Mensch-Roboter-Interaktion**

1. Auflage 2019
Dortmund/Berlin/Dresden

Diese Veröffentlichung ist der Abschlussbericht zum Projekt F 2369 „Team-Mental-Models in Mensch-Roboter-Teams: Einsatzszenarien und Stand der Technik“, Teilprojekt 2, im Auftrag der Bundesanstalt für Arbeitsschutz und Arbeitsmedizin. Die Verantwortung für den Inhalt dieser Veröffentlichung liegt bei den Autorinnen und Autoren.

Autorinnen/Autoren: Linda Onnasch, Thomas Jürgensohn
HFC Human-Factors-Consult GmbH
Köpenicker Straße 325, Haus 40, 12555 Berlin

Peter Remmers, Christoph Asmuth
Institut für Philosophie, Literatur-, Wissenschafts-
und Technikgeschichte
Straße des 17. Juni 135, 10623 Berlin

Titelbild: Phuchit/iStock.com

Umschlaggestaltung: Susanne Graul
Bundesanstalt für Arbeitsschutz und Arbeitsmedizin

Herausgeber: Bundesanstalt für Arbeitsschutz und Arbeitsmedizin (BAuA)
Friedrich-Henkel-Weg 1 – 25, 44149 Dortmund
Postanschrift: Postfach 17 02 02, 44061 Dortmund
Telefon 0231 9071-2071
Telefax 0231 9071-2070
E-Mail info-zentrum@buaa.bund.de
Internet www.buaa.de

Berlin: Nöldnerstraße 40 – 42, 10317 Berlin
Telefon 030 51548-0
Telefax 030 51548-4170

Dresden: Fabricestraße 8, 01099 Dresden
Telefon 0351 5639-50
Telefax 0351 5639-5210

Die Inhalte der Publikation wurden mit größter Sorgfalt erstellt und entsprechen dem aktuellen Stand der Wissenschaft. Für die Richtigkeit, Vollständigkeit und Aktualität der Inhalte übernimmt die BAuA jedoch keine Gewähr.

Nachdruck und sonstige Wiedergabe sowie Veröffentlichung, auch auszugsweise, nur mit vorheriger Zustimmung der Bundesanstalt für Arbeitsschutz und Arbeitsmedizin.



doi:10.21934/buaa:bericht20190128 (online)

www.buaa.de/dok/8813082

Inhaltsverzeichnis

	Seite
Kurzreferat	5
Abstract	6
1 Einleitung	7
1.1 Ausgangslage und Projektziel	7
1.2 Projektvorgehen	8
2 MRI im ethisch-moralischen Diskurs	9
2.1 Ethisch-moralische Entscheidungen von Robotern / <i>Artificial Moral Agents</i>	9
2.2 Die Zuordnung von Verantwortung für Roboter-Handlungen	17
2.3 Ethische Implikationen der Mensch-Roboter-Interaktion	19
2.4 Zwischenfazit	22
3 Kritische Aspekte der MRI aus ethisch-sozialer Perspektive	24
3.1 Ethisch-soziale Qualität der Arbeit in der MRI	26
3.1.1 Entscheidungsautorität: Führende Funktionen von Robotern	28
3.1.2 Kommunikation: Die soziale Dimension der Sprache	31
3.2 Ethisch-soziale Implikationen humanoider und zoomorpher Projektionen	32
3.2.1 Verantwortungsdiffusion	34
3.2.2 Affektive Bindungen des Menschen an den Roboter	35
4 Analyse von MRI-Studien in Bezug auf ethische und soziale Aspekte	37
4.1 Spezifizierung ethischer Aspekte	37
4.2 Studienauswahl	39
4.3 Systematisierung der Einflussgrößen anhand der MRI-Taxonomie	45
4.3.1 Die MRI-Taxonomie	45
4.3.2 Anwendung der Taxonomie auf unabhängige Variablen der Studien	46
4.3.3 Erweiterung der MRI-Taxonomie	48
4.4 Einfluss der Interaktionsrolle auf ethisch-soziale Aspekte	49
4.5 Einfluss der Autonomie auf ethisch-soziale Aspekte	50
4.6 Einfluss der Morphologie auf ethisch-soziale Aspekte	51
4.6.1 Gestaltmorphologie	51
4.6.2 Bewegungsmorphologie	52
4.6.3 Kommunikationsmorphologie	53
4.6.4 Kontextmorphologie	55
4.7 Zusammenfassung	56
5 Empfehlungen für eine ethisch-sozial angemessene MRI	58
5.1 Einfluss von Autonomiegrad und Interaktionsrolle auf Akzeptanz und Verantwortung	58
5.2 Einfluss der Morphologie auf Verantwortungsdiffusion und affektive Bindung	59
5.3 Empfehlungen zur industriellen MRI-Gestaltung	60

6	Fazit: Abwägung der ethisch-sozialen Implikationen in der MRI	62
	Literaturverzeichnis	66
	Abbildungsverzeichnis	72
	Tabellenverzeichnis	73

Ethische und soziologische Aspekte der Mensch-Roboter-Interaktion

Kurzreferat

Ziel des Projekts war die Herausarbeitung von ethischen und sozialen Aspekten der Mensch-Roboter-Interaktion, die vor allem im industriellen Kontext relevante Größen darstellen. Dazu wurde in einem ersten Projektschritt eine umfassende Literaturrecherche zur MRI im ethisch-moralischen Diskurs vorgenommen. Im Fokus standen Fragen des ethisch-moralischen Status autonomer Entscheidungen von Robotern, der Verantwortung und der Technikfolgen der MRI. Darauf aufbauend wurden im zweiten Arbeitsschritt kritische Aspekte der MRI aus ethisch-sozialer Perspektive ermittelt und deren Relevanz anhand publizierter experimenteller Studien untersucht. Zur Identifikation der einflussnehmenden Variablen in der Mensch-Roboter-Zusammenarbeit wurde die Taxonomie von Onnasch, Maier und Jürgensohn (2016) auf die ausgewählten Studien angewandt. Insgesamt wurden drei Einflussgrößen auf ethisch-soziale Aspekte der MRI durch die Analyse identifiziert: die Interaktionsrolle des Menschen, der Autonomiegrad des Roboters sowie die Morphologie des Roboters in Bezug zu Kontext, Gestalt, Kommunikation und Bewegung. Diese haben in der Ausgestaltung einen wesentlichen Einfluss auf die affektive Komponente der MRI, auf die wahrgenommene Verantwortung in der Aufgabenbearbeitung und die Akzeptanz des „Kollegen Roboters“ in verschiedenen Interaktionsrollen. Die Ergebnisse der Analyse bilden die Grundlage einer Ableitung von Gestaltungshinweisen für die industrielle MRI nach ethisch-sozialen Kriterien.

Schlagwörter:

Mensch-Roboter-Interaktion, RoboEthik, kollaborative Roboter, Ethik, Vertrauen, Nutzerakzeptanz, Verantwortung

Socio-ethical aspects of human-robot-interaction

Abstract

The objective of the presented project was to identify socio-ethical aspects of human-robot-interaction that are especially relevant in the industrial domain. Therefore, the project started with an extensive literature review with regard to the moral-ethical status of decisions made by robots, questions of responsibility, and the socio-ethical impact of the introduction of robots to the working environment.

Following this review, socio-ethical aspects were identified and their relevance was evaluated based on experimental HRI-studies published in scientific peer-reviewed journals. Relevant aspects were classified according to the HRI-taxonomy provided by Onnasch, Maier & Jürgensohn in 2016. Following this approach, three essential variables were identified that impact emotional aspects, the perceived responsibility, and the user acceptance in the industrial HRI: the hierarchical role of robot and human in the interaction, level of robot autonomy, and robot morphology in terms of shape, context, communication, and motion.

The results form the basis of socio-ethical design guidelines for industrial HRI.

Key words:

Human-robot-interaction, RoboEthics, collaborative robots, ethics, trust, user acceptance, responsibility

1 Einleitung

1.1 Ausgangslage und Projektziel

Sowohl in der Industrie als auch im Servicebereich ist ein verstärkter Einsatz von Robotern in unmittelbarer Interaktion mit dem Menschen zu beobachten. Diese Tendenz innerhalb der Robotik hin zu Assistenzrobotern, kollaborativen sowie kooperativen Interaktionsformen wird auch in Zukunft weiter zunehmen (Hooper, Spencer, & Slok, 2016). Diese neue Form der Zusammenarbeit tangiert einerseits technische Aspekte, aber auch in großem Maße ethische, rechtliche und soziologische Aspekte (ELSI-Aspekte).

Ethische und soziale Aspekte der Mensch-Roboter-Interaktion (MRI) zeigen sich zunächst im Rahmen der Verantwortungsproblematik. Ganz- oder teilautonom agierende Roboter können Schäden verursachen, sind aber offensichtlich nicht auf dieselbe Weise verantwortlich wie natürliche oder juristische Personen. Darüber hinaus kann in bestimmten Interaktions- und Teamformen die Verantwortung für die Handlungsergebnisse unter Umständen nicht mehr eindeutig zugeordnet werden. Das gilt insbesondere für autonome und lernende Robotersysteme, in denen nicht mehr nur Entwurf, Konstruktion und Programmierung, sondern auch die jeweilige Lerngeschichte und damit auch der Verlauf der Interaktion selbst die Handlungen bestimmen (Marino & Tamburrini, 2006).

In sozialer Hinsicht und im gesellschaftlichen Rahmen wirft eine zunehmende Mensch-Roboter-Interaktion Fragen hinsichtlich der Eingriffe in die bestehende Lebenswelt und zu den damit einhergehenden Veränderungen auf. Der Umgang mit Robotern wird in absehbarer Zeit nicht nur zu einer Anpassung der Roboter an den Menschen führen, sondern auch umgekehrt zu einer Anpassung der interagierenden Menschen an den Roboter. In diesem Prozess kommt es zu ganz unterschiedlichen Konstellationen, die daraufhin zu beurteilen sind, ob die erforderlichen und sich abzeichnenden Anpassungsleistungen der Menschen in ethischer und sozialer Hinsicht akzeptabel oder wünschenswert sind.

Diesen Fragestellungen nimmt sich die Ethik der Robotik (oder *Roboethik*) an, eine Strömung der angewandten Ethik (Veruggio, 2005). Im Vordergrund stehen die Probleme der Verantwortungszuschreibung an der Schnittstelle zwischen Handlungstheorie und Technikphilosophie (Rammert & Schulz-Schaeffer, 2002) sowie die Zusammenhänge zwischen der Zuschreibung intentionaler Handlungen und ethisch-moralischen Wertekontexten (Hindriks, 2011). Darüber hinaus geht es um eine Technikfolgenabschätzung der Robotik (Christaller et al., 2001), genauer der Folgen, die sich aus bestimmten Arten der MRI ergeben.

Ziel des Projekts ist es, ethische und soziale Aspekte der Mensch-Roboter-Interaktion herauszuarbeiten, die vor allem auch bei der Anwendung im industriellen Kontext relevante Größen darstellen. Basierend auf Erkenntnissen aus theoretischen Analysen sollen konkrete Gestaltungsempfehlungen der MRI nach ethisch-sozialen Kriterien entwickelt werden.

1.2 Projektvorgehen

Das erste Arbeitspaket stellt eine umfassende Literaturrecherche zur MRI im ethisch-moralischen Diskurs dar. Hier geht es in erster Linie um den ethisch-moralischen Status autonomer Entscheidungen von Robotern, um Fragen der Verantwortung und um die Technikfolgen der Robotik. Darauf aufbauend werden im zweiten Arbeitspaket kritische Aspekte der MRI aus ethisch-sozialer Perspektive theoretisch ermittelt. Im Vordergrund stehen dabei Aspekte der Entscheidungsautorität im Team, der sozialen Dimension der Kommunikation im Arbeitskontext, Probleme der Verantwortungszuordnung und der affektiven Bindung des Menschen an Roboter. Die Relevanz der Überlegungen wird anhand publizierter experimenteller Studien untersucht. Zur Identifikation der einflussnehmenden Variablen wird die MRI-Taxonomie von Onnasch, Maier und Jürgensohn (2016) auf die ausgewählten Studien angewandt. So können Einflussgrößen auf ethisch-soziale Aspekte systematisiert und konkrete Wirkzusammenhänge erkannt werden. Dies bildet die Grundlage zur Ableitung von Gestaltungshinweisen für die industrielle MRI nach ethisch-sozialen Kriterien.

2 MRI im ethisch-moralischen Diskurs

Beiträge zur Ethik der Robotik zeichnen sich – wie in der zeitgenössischen Ethik üblich – durch ein analytisches Interesse aus. Die Hauptziele der ethischen Forschung bestehen in der Identifikation von Problemen, im Aufzeigen von Optionen und in der Begründung von Bewertungen. Im Gegensatz zu anderen disziplinären Ansätzen sind die Problemstellungen nicht immer unmittelbar gegeben; das gilt insbesondere für prospektive Technikbewertungen, wie sie für das Gebiet der Robotik angezeigt sind. In vielen Beiträgen findet man daher zunächst anschauliche Beispiele, mit denen die ethische Relevanz der Robotik erst deutlich gemacht werden muss. Auch geht es zumeist nicht um abschließende normative Bewertungen technologischer Entwicklungen. Eindeutige ‚Lösungen‘ für ethische Probleme oder für eine ethisch verantwortliche Anwendung der Robotik kann es im engeren Sinne nicht geben, weil praktische Bewertungen und normative Entscheidungen theoretisch unterbestimmt sind. Bewertungen sind immer Ergebnisse von Abwägungen, die nicht im wissenschaftlichen Sinne verifizierbar sind – bestenfalls können bestimmte Szenarien als ethisch nicht vertretbar einhellig abgelehnt werden. Grundsätzlich hängen ethische Bewertungen und Entscheidungen aber von gesellschaftlichen Konsensfindungen und inhaltlichen Präferenzen ab, über die aus einer rein theoretischen Perspektive nicht abschließend geurteilt werden kann. Aus diesen Gründen besteht das Ziel ethischer Beiträge zu Technologien wie der Robotik darin, eine Sensibilität für ethische Implikationen zu vermitteln und das Feld der Handlungsoptionen abzustecken. Für die schlussendlichen Entscheidungen für oder gegen eine bestimmte technische Anwendung verweisen Ethiker schließlich an die Verantwortungsträger in Technik, Politik und Gesellschaft. Ethische Untersuchungen und Gutachten erarbeiten hierfür die konzeptionellen Grundlagen, um möglichst reflektierte und ethisch verantwortungsvolle Entscheidungen zu ermöglichen.

2.1 Ethisch-moralische Entscheidungen von Robotern / *Artificial Moral Agents*

Eine insbesondere in der englisch-sprachigen Debatte verbreitete Tendenz ethischer Auseinandersetzungen bezieht sich auf die spezifischen ethisch-moralischen Fragen und Probleme, die durch die Autonomie von Robotern aufgeworfen werden.¹ Unter den Bezeichnungen ‚machine ethics‘ oder auch ‚machine morality‘ wird nach den Bedingungen für eine Realisierung von Robotern als ‚künstlichen moralischen Akteuren‘ (artificial moral agents, AMA) gefragt (Wallach & Allen, 2009; Anderson & Anderson, 2011; Gunkel, 2012). In diesem Forschungsfeld können drei grundsätzliche Problemkomplexe identifiziert werden: 1. Die Ersetzung oder Verbesserung ethisch relevanter Entscheidungen durch Roboter, 2. die Eröffnung ethisch relevanter Ent-

¹ Die begriffliche Unterscheidung zwischen *Moral* und *Ethik* ist nicht einheitlich definiert. Mit dem Begriff ‚moralisch‘ werden im Folgenden Entscheidungen und Handlungen charakterisiert, die auf der Grundlage allgemein geteilter und begründbarer Regeln des Zusammenlebens bewertet werden (z. B. als ‚gut‘ oder ‚verwerflich‘). Mit dem Begriff ‚ethisch‘ nehmen wir dagegen Bezug auf Vorgänge, die die Regeln und Praktiken des Zusammenlebens selbst verändern. Ethische Werte beziehen sich im Unterschied zu moralischen Bewertungen auf wünschenswerte oder unerwünschte Veränderungen des Zusammenlebens, der Lebenswelt und der Gesellschaft.

scheidungsspielräume durch Roboter und 3. die Idee einer ethischen Reflexionsfähigkeit von Robotern.

1. Wenn Roboter komplexe menschliche Tätigkeiten übernehmen, dann müssen Entscheidungen, die mit diesen Tätigkeiten notwendig verbunden sind, offensichtlich ebenfalls von Robotern vollzogen werden. Die Technologien der autonomen Robotik und der intelligenten Computersteuerung zeichnen sich insofern gegenüber anderen Technologien dadurch aus, dass neben komplexen Vorgängen auch die damit verbundenen Entscheidungen automatisiert werden. Die Auswahl bestimmter Handlungsoptionen wird auf technische Systeme übertragen. Ein in diesem Sinne *autonomes* technisches System ist nicht mehr nur Mittel zur Realisierung menschlicher Absichten, sondern es legt in gewissen Grenzen selbst fest, welche Ereignisse in Gang gesetzt werden. In der ethischen Diskussion der Robotik geht es daher in erster Linie um den Einsatz von Robotern in sensiblen Bereichen, also um eine Klasse von Tätigkeiten und Handlungen, mit denen grundsätzliche ethische und moralische Entscheidungen verbunden sind. Beispielhafte Einsatzbereiche der Robotik, für die entsprechende Fragen diskutiert werden, sind etwa Militär, Medizin und soziale Robotik. In diesen Bereichen müssen Entscheidungen mit z. T. dramatischer ethischer Reichweite getroffen werden. Doch während Menschen über alltägliche oder professionelle Kompetenzen verfügen, durch die sie ethische Entscheidungen mit großer Selbstverständlichkeit und eventuell ohne tiefere Überlegungen treffen, müssen Robotern entsprechende Entscheidungsfähigkeiten erst ‚beigebracht‘ werden. Hier knüpfen sich nun grundsätzliche Fragen nach der Gestaltung dieser Fähigkeit an: Welche ethischen Prinzipien autonomer Entscheidungen für eine Handlungsoption können technisch realisiert werden? Wie können alltägliche oder professionelle ethische Entscheidungen von Menschen annähernd getreu in einem technischen System abgebildet werden? Und: Sollten Roboter überhaupt genauso entscheiden, wie Menschen an ihrer Stelle entscheiden würden? Könnten Roboter nicht – als technische Systeme – prinzipiell *besser* als Menschen entscheiden? Wie können Entscheidungen von technischen Systemen unsere moralischen und ethischen Ansprüche an sie erfüllen? – In den letzteren drei Fragen klingt bereits das Bedürfnis nach einer technischen Erweiterung der Leistungsfähigkeit an, das sich natürlich auch auf die Entscheidungsfähigkeit ausdehnt.

Prinzipiell gibt es verschiedene Szenarien für die Implementierung einer Entscheidungsfähigkeit in einem *artificial moral agent*. So könnte ein Roboter den Vorschriften einer ethischen Theorie folgen. Üblicherweise werden in diesem Zusammenhang die ‚klassischen‘ ethischen Entscheidungsprinzipien vorgestellt und diskutiert, die den einschlägigen Theorien der philosophischen Ethik korrespondieren: *Konsequentialismus* bzw. *Utilitarismus* und *deontologische Ethik*. Utilitaristische Kriterien bewerten Handlungsoptionen nach dem zu erwartenden Gesamtnutzen aus allen beteiligten Werten und Interessen. Diejenige Handlung, die den höchsten Gesamtnutzen verspricht, ist ethisch am besten zu bewerten. Im äußersten Falle können im utilitaristischen Paradigma auch Menschenleben gegeneinander aufgerechnet werden. Demgegenüber setzen deontologische Theorien unveräußerliche Grundwerte fest, die nicht mit anderen Werten ‚verrechnet‘ werden dürfen. Beispielsweise sind Men-

schenleben unter allen Umständen zu schützen, auch wenn der Gesamtnutzen dadurch nicht maximiert wird (Tzafestas, 2016).

Welcher der konkurrierenden ethischen Theorien sollte ein Roboter folgen? Bei der Beschreibung der menschlichen Entscheidungsfähigkeit ist realistisch von einer Mischform verschiedener Theorien auszugehen, wobei die genaue Gewichtung allerdings umstritten ist. Ein Vorschlag zur Festlegung der Entscheidungskriterien besteht in einem moralischen Turing-Test, der auf eine Ununterscheidbarkeit zwischen moralischem Verhalten von Robotern und Menschen abzielt (Allen, Varner, & Zinser, 2000), gegen dessen Konzeption allerdings zuletzt starke Argumente formuliert wurden (Arnold & Scheutz, 2016). Eine allgemein anerkannte Lösung für eine ‚programmierte Ethik‘ zeichnet sich bisher nicht ab. Vernünftig erscheint hier aber der Hinweis, dass entsprechende Entscheidungsmodelle „so weit wie möglich durch eine gesellschaftliche Konsensbildung nach ausführlicher öffentlicher Diskussion festgelegt werden sollten“ (Birnbacher & Birnbacher, 2016).

2. Ethische Relevanz erhalten neue Technologien schließlich auch dadurch, dass sie Entscheidungsspielräume überhaupt erst eröffnen, die zuvor noch nicht verfügbar waren. Denn mit technischen Mitteln wird eine Steuerung von Prozessen möglich, auf die es bisher gar keine effektiven Einflussmöglichkeiten gab. So erweitern Kraft, Präzision und Zuverlässigkeit technischer Systeme die Handlungsmöglichkeiten beträchtlich. Doch sie bringen zugleich auch neue Verantwortungsansprüche mit sich (Jonas, 2003). Ein bekanntes Beispiel, mit dem diese Forschungsfrage veranschaulicht werden kann, ist das sogenannte Trolley-Szenario, ein Dilemma, das u.a. im Zusammenhang mit autonomen Fahrzeugen häufig diskutiert wird (Birnbacher & Birnbacher, 2016): Wenn ein Kraftfahrzeugführer die Wahl hat zwischen dem Unfalltod mehrerer Fußgänger oder einer Selbsttötung – wofür sollte er sich entscheiden? An diese Frage anschließend ergeben sich vor dem Hintergrund der genannten ethischen Theorien verschiedene Ansatzpunkte. Wiegen mehrere Menschenleben schwerer als der Tod eines Einzelnen? Oder gibt es ein unbedingtes Gebot der Selbsterhaltung? In praktisch relevanten Fällen des Autofahrens konnte dieses Dilemma bisher vernachlässigt werden, denn man geht davon aus, dass die Bedenkzeit eines Menschen in einer derartigen Situation zu kurz für eine absichtliche und planmäßige Entscheidung ist. Aus diesem Grund würde ein Fahrer wohl nicht für die Folgen einer entsprechenden Entscheidung verantwortlich gemacht werden. Das Dilemma bekommt allerdings im Rahmen der Entwicklung selbstfahrender Kraftfahrzeuge neue Relevanz, denn hier kann dieses Szenario prinzipiell im Voraus bei der Programmierung des Fahrzeugs berücksichtigt werden. Es entsteht folglich ein neuer Entscheidungsspielraum, der rationalen Kriterien zugänglich wird und durch programmierte Verhaltensweisen auszufüllen ist. Wo bisher ‚blinde Flecken‘ der ethischen Bewertung lagen, erzeugen Technologien autonomer Steuerungen neue Ansprüche an eine ethisch verantwortungsvolle Gestaltung. Die ethische Relevanz der Robotik verdankt sich daher besonders auch denjenigen technischen Leistungen, die über die Möglichkeiten des Menschen hinausgehen. Beispielsweise können Roboter in Bruchteilen von Sekunden Entscheidungen treffen, prinzipiell schneller als Menschen. Oder sie können zwischen ver-

schiedenen Optionen auf der Basis einer ungleich größeren Informationsbasis auswählen.

Durch diese Möglichkeiten einer technischen Umsetzung verändern sich aber zugleich auch die Ansprüche an die ethische Reflexion und deren konkrete Umsetzung in Regeln, weil ein zuvor vernachlässigbarer Handlungsbereich für geplante Eingriffe zugänglich gemacht wird. Was also zunächst wie eine Entlastung des Menschen erscheint – indem komplexe Tätigkeiten und Entscheidungen an Roboter delegiert werden – stellt sich schließlich als zusätzliche Aufgabe heraus, die von der Robotik zu lösen ist. Es entsteht ein neuer Entscheidungsspielraum, der rationalen Kriterien zugänglich wird und durch programmierte Verhaltensweisen auszufüllen ist.

3. Anknüpfend an diese Anforderungen diskutieren einige Autoren auch die Idee einer höherstufigen moralischen Entscheidungsfähigkeit oder gar eines ethischen Reflexionsvermögens von Robotern. In diesem Zusammenhang wird die oben dargestellte einfachere Vorstellung einer ‚programmierten‘ Befolgung moralischer Regeln unterschieden von der anspruchsvolleren Idee einer ethischen Reflexionskompetenz von Robotern (Veruggio & Operto, 2008). Eine derartige Reflexionskompetenz würde verantwortungsvolle Entscheidungen und Handlungen ermöglichen – auch in Situationen, in denen die Frage nach dem ethisch-moralisch ‚Richtigen‘ umstritten oder noch unbeantwortet ist. Solche Szenarien setzen allerdings offensichtlich eine starke KI voraus, wodurch sie nicht auf zeitnahe Entwicklungen bezogen sind. Unter anderem aus diesem Grund haben entsprechende Vorstellungen Kritik auf sich gezogen. Einige Autoren melden grundsätzliche Zweifel daran, dass die diskutierten Reflexionsfähigkeiten in absehbarer Zukunft (oder jemals) realisierbar sind. Beiträge, in denen die Realisierbarkeit nicht eigens thematisiert und vorausgesetzt wird, bewegen sich Kritikern zufolge im Bereich unbegründeter und praktisch irrelevanter Spekulationen. Datteri und Tamburrini bemerken entsprechend: „To the best of our knowledge, however, the moral agentivity of robots is not a pressing ELS issue, insofar as it concerns no technologically foreseeable robotic system“ (Datteri & Tamburrini, 2013, S. 213).

Eine Motivation für die drei dargestellten Untersuchungsrichtungen erklärt sich daraus, dass damit ein unmittelbarer Zusammenhang zwischen allgemeinen philosophisch-ethischen Überlegungen und technischen Umsetzungen hergestellt wird. Auch philosophische Theorien zu den Bedingungen ethisch-moralischen Handelns spielen in dieser Diskussion eine wichtige Rolle. Konzepte wie Verantwortung, Autonomie und Person werden dann nicht mehr nur als Grundlagen menschlicher Praktiken und Ideale analysiert, sondern aus der Perspektive einer technologischen Ermöglichung. Dieser Zusammenhang eröffnet auch eine Perspektive auf eine interdisziplinäre Zusammenarbeit zwischen Philosophen und Ingenieuren (Anderson & Anderson, 2011). Denn es geht um die Frage, wie sich moralische Regeln oder ethische Reflexionsvermögen technisch konkret implementieren lassen, wie sich folglich Entscheidungsprozeduren bzw. Algorithmen für ethisch-moralisches Verhalten in Robotern realisieren lassen (Pereira & Saptawijaya, 2016).

Auf der anderen Seite besteht die Gefahr, dass sich die Diskussion der genannten Probleme von seinem spezifischen Bezug zur Robotik entfernt und in die Richtung

allgemeinerer ethischer Fragestellungen tendiert. Wie ist beispielsweise im Trolley-Dilemma zu entscheiden? Zwar ist die Entwicklung autonomer Kraftfahrzeuge ein konkreter Anlass, über diese Frage nachzudenken, aber die Antwort auf diese Frage ist völlig unabhängig von jeder technischen Realisierung. Mit der autonomen Technologie eröffnet sich ein Handlungsspielraum, der bei der Formulierung des Dilemmas in der Theorie bereits vorausgesetzt ist und insofern keine spezifische Bedeutung für theoretische Überlegungen hat. So wird beispielsweise die Idee eines künstlichen moralischen Akteurs nicht in erster Linie vor dem Hintergrund praktischer Szenarien diskutiert, sondern hauptsächlich mit dem Ziel, über die Bedingungen von Ethik und Moralität im Allgemeinen nachzudenken. In Untersuchungen von dieser Art geht es nicht unbedingt um realistische Szenarien der Robotik im engeren Sinne; sie gilt nur als konkreter Spezialfall eines allgemeineren theoretischen Problems. Vielmehr besteht ein methodisches Interesse an der Kontrastierung von menschlichen und künstlichen Akteuren: Der Vergleich mit Robotern eröffnet die Perspektive auf Kriterien, die für Konzepte wie Handlung und Verantwortung im Allgemeinen gelten – unabhängig von idiosynkratischen menschlichen Praktiken. In diesem Sinne kritisiert beispielsweise P. M. Asaro Ansätze, die sich auf die Entwicklung eines „moral reasoning system“ beziehen, für ihre extreme Abstraktion und ihr hintergründiges rein theoretisches Interesse (Asaro, 2006).

Einschätzung der ethischen Entscheidungsfähigkeit von Robotern als Technologieziel

In der aktuellen Forschung wird die Implementierung einer ethischen Entscheidungsfähigkeit von Robotern von einer angemessenen Formalisierung ethischer Prinzipien abhängig gemacht. Es stellt sich allerdings die Frage, inwiefern eine solche Entscheidungsfähigkeit überhaupt wünschenswert ist. Was sollte das Ziel des Einsatzes von künstlichen ethisch-moralischen Akteuren sein?

Zur Diskussion dieser Frage sind allgemein zwei Ziele der Entwicklung von technisch-autonomen Entscheidungsträgern zu unterscheiden:²

1. Zunächst kann es um die Entwicklung von Robotern gehen, deren planmäßiger Zweck darin besteht oder wesentlich davon abhängt, in ethisch heiklen Situationen autonome Entscheidungen zu treffen. Im Folgenden wird dafür argumentiert, dass das Ziel in einer *Minimierung* von Festlegungen auf ethisch umstrittene Positionen bestehen sollte. Der Einsatz von Robotern, deren Zweck gerade darin besteht, in ethisch umstrittenen Situationen autonom zu entscheiden, ist hingegen abzulehnen.
2. Roboter, die für andere Zwecke eingesetzt werden, können dennoch in ungewollte Extremsituationen geraten, in denen eine ethisch-moralische Entscheidungsfähigkeit notwendig wird. Die Entwicklung eines autonomen Roboters muss die Möglichkeit einer solchen Situation offenbar berücksichtigen. Die Option, im Voraus gar keine Entscheidung über den Ausgang des hypotheti-

² Ähnlich wie autonome Handlungsentscheidungen ist übrigens auch die autonome Selektion von Informationen durch technische Systeme zu betrachten. Sofern auf der Basis der selektierten Informationen menschliche Entscheidungen mit ethischer Dimension getroffen werden, ergibt sich daraus eine mittelbare Einflussnahme der Technik. Hier ist zusätzlich zu den im Folgenden diskutierten Punkten ein Mindestmaß an struktureller Transparenz notwendig.

schen Unfalls zu treffen, ist hier nicht mehr gegeben. Hier kann auf dem aktuellen Stand der Forschung lediglich festgestellt werden, dass keine diskriminierenden Kriterien in technischen Entscheidungsprinzipien implementiert werden dürfen (vgl. S. 12f.).

Zu 1. Allgemein ist davon auszugehen, dass die ethisch-moralischen Konsequenzen von Roboter-Entscheidungen grundsätzlich anders bewertet werden als gleichartige Entscheidungen von Menschen. Mit Fortschritten der technischen Entwicklung geht der Anspruch an eine tiefgreifende *Verbesserung* der Verhältnisse einher. Vor dem Hintergrund dieser Verbesserungen werden die neuen Unsicherheiten und Gefahren technischer Innovationen bewertet. Wenn nun autonom entscheidende Maschinen für den Einsatz in ethisch heiklen Situationen entwickelt werden, dann müssen die Entscheidungen zu insgesamt wesentlich besseren Ergebnissen führen. Inakzeptabel wäre es dagegen, wenn Roboter genau wie Menschen (oder gar ‚schlechter‘) entscheiden würden, denn Menschen handeln häufig unmoralisch und ethisch fragwürdig. Darüber hinaus ist der Bereich ethisch-moralischer Festlegungen so sensibel, dass sich der Einsatz neuer Technologien daran messen muss: Die Toleranz für deutlich problematische Entwicklungen im ethischen Bereich ist im Allgemeinen niedriger anzusetzen als beispielsweise in Bezug auf Nachteile, die sich in rein technischer Hinsicht ergeben.

Vor diesem Hintergrund ist festzuhalten: Aus ethischer Perspektive ist der Einsatz technischer Systeme mit autonomen Entscheidungsfähigkeiten immer dann problematisch, wenn sie gezielt in Bereiche eingreifen, die normalerweise von ethisch-moralischen Konflikten geprägt sind. Technisch realisierte Entscheidungen mit ethisch-moralischer Dimension sind ausschließlich dort wünschenswert und geboten, wo von einem allgemeinen Konsens über die ethische Richtigkeit der Entscheidung ausgegangen werden kann. Der Entscheidungsspielraum für autonome Systeme dürfte sich daher weitgehend mit den Bedingungen für einen möglichst sicheren Betrieb decken, so dass sich alle autonomen Entscheidungen an unumstrittenen ethischen Ansprüchen des Menschen orientieren.

Die Diskussion einer autonomen Entscheidungsfähigkeit technischer Systeme dreht sich allerdings häufig gerade um Szenarien (wie die oben dargestellten Trolley-Beispiele), in denen ein moralischer Konsens gar nicht gegeben ist. Die Intuitionen gehen hier auseinander – eine Zuordnung in utilitaristische und deontologische Theorien kann hier bestenfalls die zugrundeliegenden Prinzipien verdeutlichen, führt aber nicht zur Einigkeit. Diese Verhältnisse ergeben sich schon dann, wenn sich Personen unsicher über die richtige Entscheidung sind; auch würden verschiedene Menschen dieselbe Situation unterschiedlich beurteilen – je nach persönlichen, gruppenspezifischen oder kulturellen Werten, Gefühlen, Traditionen etc. Auch komplexere Theorien oder anspruchsvolle Formalisierungen des moralischen *common sense* tragen hier nicht zu einem Konsens bei. In diesem Sinne gibt es keine ‚ideale‘ Entscheidung, an der sich ein technisches System orientieren kann. Nicht die objektiven Mängel menschlicher Entscheidungen, sondern die prinzipielle Unbestimmtheit einer weiten Klasse ethisch-moralischer Entscheidungen sind für diesen Mangel verantwortlich.

Wie ist mit dem Einwand umzugehen, dass technische Systeme im Allgemeinen *bessere* Entscheidungen treffen, z. B. weil sie über mehr Informationen verfügen

oder weil sie nicht von den ‚richtigen‘ Prinzipien abweichen (etwa aus emotionalen Beweggründen)? Man kann in Bezug auf konkrete Szenarien überlegen, ob die bestehende Uneinigkeit in Bezug auf ethisch-moralische Entscheidungen von Bedingungen abhängt, von denen im Rahmen einer technischen Realisierung nicht mehr auszugehen wäre. Gerade im Fall des Trolley-Beispiels ist das aber nicht der Fall: Es sind keine verbesserten oder zusätzlichen Eigenschaften eines technischen Systems denkbar, die im Trolley-Beispiel einen eindeutigen Lösungskonsens ermöglichen. Immer dann, wenn die ethisch-moralische Richtigkeit einer Entscheidung prinzipiell diskutabel ist, verletzen technisch realisierte autonome Entscheidungen folglich den Wert der Autonomie des Menschen. Sie führen zu einer *technisch-rationalen Regulierung* bzw. *Normierung* ethischer Entscheidungen, über deren Richtigkeit aus guten Gründen bisher keine Einigkeit erzielt wurde. Vielmehr gilt: In Situationen, die ethisch-moralisch nicht eindeutig bewertet werden können, ist eine Verletzung ethischer Werte wie Leben oder Gesundheit durch technische Systeme grundsätzlich immer inakzeptabel, wenn sie nicht auf menschliche oder technische Fehler zurückzuführen ist.

Technische Implementierungen ethisch kontroverser Entscheidungen führen zu Konsequenzen, die sich ähnlich wie eine ‚Verrechtlichung‘ eines ethischen Entscheidungsmodells auswirken würden. Dieser Vorgang unterscheidet sich grundsätzlich von politischen Entscheidungen mit ethischem Grundsatzcharakter, beispielsweise zur gentechnischen Forschung: Hier werden begründete (politische) Entscheidungen mit ethischer Reichweite getroffen, die allerdings kontrovers bleiben und gegen andere Auffassungen festgesetzt werden. Der Unterschied zwischen politischen Regelungen dieser Art und der Festlegung technischer Entscheidungsprozeduren (mit ‚Freiheitsgraden‘) besteht darin, dass in der Technik keine umstrittene Position, sondern ein *Entscheidungsprinzip* etabliert wird. Der Prozess, der in der Politik aufwändig für einzelne Entscheidungen in Gang gesetzt wird, würde formalisiert und in technischen Systemen für ganze Klassen von kontroversen Entscheidungen festgesetzt.

Der Verzicht auf eine Entwicklung von Robotern, deren Zweck in gezielten Entscheidungen ethischer Problemsituationen besteht, entbindet dagegen von der Notwendigkeit, an Stelle aller Betroffenen über diese ethischen Konflikte zu entscheiden. Die Autonomie des Menschen bleibt so gewahrt. Zugleich wird damit der Verlagerung von Verantwortung entgegengewirkt, die mit technischen Entscheidungen verbunden ist (vgl. Abschnitt 2.2).

Es ist im Übrigen nicht ausgeschlossen, dass sich in Zukunft ein gesellschaftlicher Konsens dahingehend entwickelt, eine Verletzung ethischer Werte durch störungsfreie Entscheidungen technischer Systeme ohne menschlichen Einfluss anzuerkennen. Die Entscheidungen eines technischen Systems könnten so zum Maßstab für das ethisch-moralisch Richtige werden. Aber gegenwärtig ist das nicht der Fall – im Gegenteil: Es ist mit einer Skepsis gegenüber der Technisierung ethisch-moralischer Entscheidungen zu rechnen (vermittelt nicht zuletzt durch negative Szenarien in der Science Fiction).

Zu 2. In der Debatte um die ethische Dimension autonomer Entscheidungen von Robotern werden gelegentlich Szenarien diskutiert, in denen Roboter gefährdete Personen nach bestimmten Maßstäben bewerten. Beispielsweise könnte ein autonomes Fahrzeug in einer Gefahrensituation nicht nur die Anzahl der gefährdeten Personen,

sondern auch deren Alter oder deren sozialen Hintergrund einbeziehen. Möglich werden derartige Szenarien insbesondere durch die zunehmende informationelle Vernetzung der Fahrzeuge. Einerseits liegt hier der Gedanke nahe, dass jegliche Diskriminierungen nach Gruppenzugehörigkeiten aus ethischer Perspektive abzulehnen sind. Andererseits werden dadurch an die Technologie wiederum wesentlich höhere Maßstäbe herangetragen, als es bei der Beurteilung von vergleichbarem menschlichem Verhalten gerechtfertigt wäre. Während Menschen besonders auch aus emotionaler Intuition heraus Entscheidungen treffen (für die sie dann allerdings in Extremsituationen auch nicht verantwortlich gemacht werden), können autonome Maschinen gerade durch den konsequenten und ausnahmslosen Ausschluss jeder Diskriminierung als ‚hart‘ oder emotional ‚kalt‘ erscheinen (Bsp. höhere Bewertung einer Gruppe von Rentnern vs. einzelnes Kind).

Aus ethischer Perspektive ist in diesen Szenarien eine Zufallsentscheidung jeder eigenschaftsbezogenen Bewertung vorzuziehen. Denn bei allgemeiner Bekanntheit von diskriminierenden Bewertungskriterien würde das Verkehrsrisiko für Personen mit entsprechenden Eigenschaften ansteigen. Selbst bei ausbleibenden Unfällen würde die Lebensqualität folglich mit bestimmten Eigenschaften korrelieren (Stichwort Sicherheitsgefühl).

In spezifischen MRI-Szenarien besteht eine allgemeine Schwierigkeit darin, den schmalen Grat zwischen Diskriminierung und sensibler Berücksichtigung von individuellen bzw. gruppenspezifischen Unterschieden in interagierenden Robotern umzusetzen. Denn es kann in einigen Szenarien durchaus sinnvoll sein, dass ein Roboter beispielsweise zwischen männlichen und weiblichen oder zwischen jungen und alten Interaktionsteilnehmern differenziert und sein Verhalten entsprechend anpasst. Hier sind Kriterien gefragt, die ein sensibles Verhalten des Roboters regeln und die eine Abwägung verschiedener Werte widerspiegeln.

Zusammenfassend kann festgehalten werden: Implementierungen für autonome Entscheidungen in Extremfällen dieser Art sollten aus ethischer Perspektive also folgenden Kriterien genügen:

- a) Es müssen Sicherheitsvorkehrungen getroffen werden, die den Eintritt eines ethischen Dilemmas im Voraus vermeiden. Die Wahrscheinlichkeit einer notwendigen autonomen Entscheidung mit ethischen Konsequenzen muss technisch minimiert werden (Birnbacher & Birnbacher, 2016).
- b) Entscheidungsprinzipien autonomer Technologien dürfen nicht zu einer generellen Verschlechterung der Systemsicherheit führen, beispielsweise indem der Anreiz zum flächendeckenden Einsatz einer statistisch sichereren Technik ausbleibt.
- c) Die potentiellen Opfer eines ethischen Dilemmas dürfen nicht nach diskriminierenden Kriterien bewertet werden. Dies gilt insbesondere für utilitaristische Bewertungsmodelle.

2.2 Die Zuordnung von Verantwortung für Roboter-Handlungen

In der Frage nach der Verantwortung für die Folgen des Einsatzes von Robotern treffen sich rechtliche und ethische Probleme. Während Verantwortung im juristischen Sinne als Haftung zu verstehen ist (Hilgendorf, 2012), betrifft Verantwortung im ethisch-moralischen Sinne Zuschreibungen wie Schuld, böse oder gute Absichten. Die oben genannten Anforderungen an ethische Entscheidungen von Robotern weisen bereits auf Veränderungen in der Zuordnung von Handlungen, Absichten und Verantwortung hin. Die im menschlichen Falle enge Verknüpfung der ursächlichen Handlung mit dem Verantwortungsträger wird bei Roboter-Handlungen gelöst. Sowohl für den rechtlichen wie auch den ethisch-moralischen Bereich stellt sich daher zunächst dasselbe Problem: Roboter führen zwar Handlungen aus, aber sie sind offenbar keine Handlungssubjekte im üblichen Sinne – weder im rechtlichen, noch im ethisch-moralischen Sinne. Es scheint sich hier eine unbefriedigende ‚Leerstelle‘ im Verantwortungszusammenhang zu ergeben. Wie ist mit dieser Situation umzugehen?

- a) Einerseits wird die Frage diskutiert, ob und wie die Verknüpfung von dem Roboter als Handlungssubjekt und einer verantwortungstragenden Rolle dadurch wieder hergestellt werden könnte, dass Roboter selbst als Verantwortungsträger gelten können. Hier steht also der Roboter selbst als mögliches Verantwortungssubjekt im Mittelpunkt der Betrachtung. Die Verantwortungsproblematik verleitet einige Roboter-Ethiker also zu der Frage: „Unter welchen Bedingungen können Roboter für ihre Handlungen verantwortlich sein?“ (Floridi & Sanders, 2004; Hilgendorf, 2012; Neuhäuser, 2014). An diese Frage schließen auch Untersuchungen dazu an, ob Robotern unter bestimmten Umständen Rechte und Pflichten zugesprochen werden sollten. Die Antworten auf derartige Fragestellungen fallen allerdings weitgehend negativ aus. Roboter werden zwar durchaus als ethisch-moralische Akteure beschrieben, sofern ihre Handlungen entsprechende Implikationen haben (Sullins, 2006). Damit ist aber keine Festlegung auf ethisch-moralische Verantwortungsfähigkeit verbunden. Das Konzept der Verantwortung setzt ein Verständnis und eine Empfänglichkeit für moralische Gründe, Verbote und Aufforderungen voraus. Es bildet die Grundlage für Lob, Tadel und Strafe und ist eng verbunden mit moralischen Gefühlen wie Empörung, Übelnehmen und Schuld. Die Zuschreibung von ethisch-moralischer Verantwortung ist nur sinnvoll in einem Kontext praktischer Rationalität. Moralisch verantwortlich ist eine Person, wenn sie Handlungsgründe verstehen, reflektieren und in Handlungen wirksam machen kann. Über derartige Fähigkeiten verfügen Roboter offensichtlich nicht. Künstliche Akteure können entsprechende Erwägungen offenbar nicht im eigentlichen Sinne ‚verstehen‘ – zumindest nicht auf dem aktuellen Stand der Entwicklung und in absehbarer Zeit. Dennoch führen sie Handlungen aus, die ethisch-moralischen Wert haben oder zu entsprechenden Konsequenzen führen. Hier zeigt sich: Ebenso wie die oben dargestellte Diskussion der Entscheidungsfähigkeit verdanken sich Fragen nach einer ethisch-moralischen Verantwortungsfähigkeit von Robotern häufig einer theoretischen Motivation. Es geht dann nicht um realistische Szenarien mit ethischen Implikationen, sondern vielmehr um die Bestimmung allgemeiner Bedingungen für Verantwortlichkeit. In diesem Zusammenhang wird kein spezifisches Problem der Robotik diskutiert.

Darüber hinaus ist die Vorstellung eines verantwortlichen Roboters auch auf mythische und fiktionale Vorstellungen einer starken KI zurückgeführt worden, die unser Bild von Robotern – trotz rationaler technischer Einstellungen – prägen (Mudry, Degallier, & Billard, 2008; Kehl & Coenen, 2016). An dieser Fragestellung der Roboter-Ethik zeigen sich deutliche Nachwirkungen unserer eingespielten Praxis zwischenmenschlicher Interaktion sowie auch kultureller Einflüsse, wie sie insbesondere der Science Fiction entstammen.

- b) Da Roboter selbst keine Verantwortungssubjekte sind, ist eine Art ‚Verschiebung‘ in der Verantwortungspraxis zu erwarten: Nicht der Roboter, der eine Handlung unmittelbar verursacht, ist verantwortlich zu machen, sondern vielmehr diejenigen im Kausalzusammenhang entfernter stehenden Personen, die die Handlungen des Roboters ermöglicht und maßgeblich bestimmt haben. Verantwortung kann so beispielsweise Personen zugeschrieben werden, die Entscheidungen zur Verwendung bestimmter Roboter fällen. Wenn z. B. vernünftigerweise zu erwarten ist, dass Roboter mit ihren Handlungen gegen rechtliche Vorschriften oder ethische Grundsätze verstoßen, dann ist unter Umständen diejenige Personengruppe verantwortlich zu machen, die den Einsatz dieser Roboter veranlasst hat. Hier ergibt sich die Schwierigkeit, die für eine Roboter-Handlung relevanten Ursachen zu identifizieren, die insbesondere im Kontext von Haftungsfragen bereits ausführlich untersucht wird (Beck, 2009). Unabhängig von dieser Schwierigkeit unterscheiden sich Situationen wie diese aber grundsätzlich von der üblichen zwischenmenschlichen Praxis, in der ein Handlungssubjekt im Normalfall für seine jeweils einzelnen Handlungen zumindest mitverantwortlich ist. In Mensch-Roboter-Interaktionen ist dagegen jegliche Verantwortung des unmittelbar agierenden Roboters in allen Fällen ausgeschlossen.
- c) Eine weitere naheliegende Lösung für das Verantwortungsproblem besteht darin, jegliche Zuschreibung von ethisch-moralischer Verantwortung unter bestimmten Umständen gänzlich auszuschließen. Sofern die Handlungen eines autonomen und lernfähigen Roboters nicht auf Handlungen von Programmierern, Konstrukteuren und Anwendern zurückzuführen sind, handelte es sich dann um einen Unfall, für den möglicherweise *niemand* ethisch-moralisch verantwortlich ist (Neuhäuser, 2014). Eine entsprechende Lösung für den ethisch-moralischen Bereich zeichnet sich etwa im Vergleich der Robotik mit der Tierethik ab, denn auch Tiere werden gewöhnlich nicht ethisch-moralisch verantwortlich für von ihnen verursachte Schäden gemacht (Birnbacher, 2013; Mudry et al., 2008). Dagegen können Haftungsfragen zugleich rechtlich klar geregelt sein. Dieser Lösungsvorschlag ist insofern beachtenswert, weil hier ein wichtiger Unterschied zwischen rechtlichen und ethischen Verantwortungsfragen zum Tragen kommt: Während die Klärung von Haftungsfragen unter einem gewissen Handlungsdruck steht, ergibt sich nicht unbedingt ein schwerwiegender Mangel, wenn ethisch-moralische Verantwortungszuschreibungen ausbleiben. Haftung im rechtlichen Sinne sollte beim Einsatz von Technik möglichst lückenlos zugeordnet werden können, ethisch-moralische Verantwortung hingegen nicht unbedingt.
- d) Schließlich gehen Mensch-Roboter-Interaktionen grundsätzlich mit *Attributionen* von Handlungsabsichten einher. Es gibt bestimmte Erwartungen von

Menschen an Roboter; in kooperativen und kollaborativen Interaktionen müssen Absichten und Handlungsziele unterstellt werden – nur so kann eine intuitive Interaktion gelingen (Bauer, Wollherr, & Buss, 2008; Sebanz & Knoblich, 2009). Nun kann dieses interaktive Verhältnis mit einer gewissen Selbstverständlichkeit zu einer Verantwortungszuschreibung führen, die den Roboter selbst im Blick hat. Dabei geht es nicht unbedingt um explizite und bewusste Zuschreibungen, sondern auch um ‚moralisch-ethische‘ Einstellungen und Gefühle, beispielsweise Übelnehmen, Ärger und Enttäuschung. Akzeptiert man die beiden zuvor dargestellten Positionen – dass Roboter selbst nicht ethisch-moralisch verantwortlich sein können und dass unter bestimmten Umständen auch niemand sonst verantwortlich für die Handlungen des Roboters sein kann –, dann müssen Zuschreibungen dieser Art als falsch beurteilt werden. Zugleich beruhen sie auf einer eingespielten Praxis und entsprechen der zwischenmenschlichen Gewohnheit, so dass sich hier ein gewisser Widerspruch zur bekannten Lebenswelt entwickeln könnte. Aus dieser Situation könnte mithin ein Konflikt entstehen, indem psychologische Neigungen (bis hin zu Bedürfnissen und ‚Zwängen‘) zur Verantwortungszuschreibung mit der rationalen Beurteilung der Sachlage konkurrieren. Dieser Konflikt wurde in Bezug auf die Verantwortungsproblematik bisher allerdings in der Forschungsliteratur noch nicht ausführlich untersucht. Entsprechende Probleme können insbesondere im Zusammenhang mit allgemeineren Tendenzen der Vermenschlichung von Robotern gesehen werden (vgl. Kapitel 2.3).

Einschätzung von Verantwortungsfragen

Verantwortung im ethisch-moralischen Sinne stellt für die Robotik aus objektiver Perspektive kein tiefgreifendes Problem dar: Obwohl Roboter handeln und entscheiden, sind sie keine Träger von Verantwortung im ethisch-moralischen Sinne. Wenn Roboter Schäden verursachen, dann sind sie – wie andere technische Artefakte – keine Teilnehmer der menschlichen ethisch-moralischen Praxis. Roboter tragen keine moralische *Schuld*. Während rechtliche Verantwortung (Haftung) den entsprechenden beteiligten Personen zuzuordnen ist, kann unter bestimmten Umständen niemand moralisch verantwortlich sein. Wie oben bereits dargelegt, haben autonome Roboter in dieser Hinsicht einen ähnlichen ethischen Status wie Tiere.

Diese ‚objektive‘ Perspektive ist aber nicht alleine maßgeblich. In der MRI werden aus der Perspektive der interagierenden Personen grundsätzlich Handlungskontexte und entsprechende ethisch-soziale Implikationen zugeschrieben. Auch wenn Roboter aus theoretischer Perspektive nicht verantwortlich *sind*, so werden sie im Rahmen der Zuschreibung intentionaler Handlungen durchaus verantwortlich *gemacht* – auch wenn das irrational erscheinen mag. Es stellt sich die grundsätzliche Frage, wie mit der ‚Doppelnatur‘ eines nur simuliert handelnden Interaktionspartners angemessen umzugehen ist (siehe dazu erste Überlegungen in Kapitel 2.3).

2.3 Ethische Implikationen der Mensch-Roboter-Interaktion

Ein anderer Ansatz in der Roboter-Ethik knüpft enger an Methoden und Fragestellung der Technikfolgen-Abschätzung an, wie sie im Rahmen der Technikethik etabliert sind. Die Robotik wird hier als besonderer Gegenstand der angewandten Ethik ähnlich wie andere relevante Technologien behandelt. Mit dieser Ausrichtung wird

nicht nur der Einsatz von Robotern in Bereichen thematisiert, in denen sensible ethische Entscheidungen getroffen werden. Vielmehr geht es in erster Linie um diejenigen ethischen und sozialen Probleme, die durch den Einsatz von Robotern überhaupt entstehen. In diesem Kontext sind auch Szenarien eingeschlossen, in denen Handlungen ohne ethischen Entscheidungsdruck vollzogen werden. Im Vergleich zur Eigenschaft der Autonomie sind in diesem Zusammenhang auch andere Kategorien der MRI relevant wie z. B. humanoide bzw. zoomorphe Gestaltung oder Interaktionsform. Es geht also um den ethisch richtigen Umgang mit der Technologie – sowohl auf Seiten der Entwickler und Hersteller, als auch auf Seiten der Anwender. Allgemein formuliert könnte die jeweils zugrundeliegende Fragestellung etwa lauten: Welche Auswirkungen hat die Mensch-Roboter-Interaktion auf gegebene Praktiken, Werte und Vorstellungen? Welche Werte könnten durch eine konkrete Ausgestaltung der MRI verletzt werden und wie wäre dem entgegenzuwirken? Sollte eine bestimmte Robotik-Technologie aus ethischen Gründen nicht eingesetzt oder gar nicht erst entwickelt werden? Wie können ethische Implikationen gezielt in der Entwicklung und im Einsatz von Robotern gestaltet werden?

Allerdings wird die Robotik aus dieser Perspektive nur selten als allgemeine und einheitliche Technologiesdisziplin thematisiert, vielmehr konzentrieren sich die Beiträge zumeist auf die spezifischen Implikationen in bestimmten Anwendungsbereichen. Intensive Debatten zum Einsatz der Robotik finden besonders in Bezug auf die folgenden Bereiche statt:

- Militär (z. B. ferngesteuerte Drohnen oder Roboter-Soldaten)
- Kranken- oder Altenpflege
- Therapie
- Medizinische Eingriffe (z. B. Chirurgie)
- Service-Roboter (z. B. Staubsauger oder Rasenmäher)

Die jeweiligen ethischen Probleme und Lösungsansätze können nur im Rahmen von bereichsspezifischen Untersuchungen erörtert werden, zumal die praktischen Kontexte und die ethischen Aspekte jeweils sehr unterschiedlich ausfallen. Abgesehen von der in Abschnitt 2.1 dargestellten Diskussion um autonome Entscheidungen von Robotern gibt es bisher nur vergleichsweise wenige Beiträge, in denen die ethisch-sozialen Implikationen der MRI im Allgemeinen thematisiert werden. Auch Fragen der Zusammenarbeit im Team werden in diesen bereichsspezifischen Untersuchungen kaum eigens thematisiert, denn es geht fast ausschließlich um das Verhältnis von Robotern zu denjenigen Personengruppen, auf die sich die ‚Handlungen‘ der Roboter beziehen, also beispielsweise auf Kriegsteilnehmer, pflegebedürftige Menschen oder Patienten. Die Arbeitsverhältnisse zu menschlichen Mitarbeitern bilden dagegen einen besonderen Fall und sind eigens zu untersuchen.

Ethische Implikationen anthropomorpher Gestaltung

In der Diskussion um die Technikfolgen der Robotik zeichnet sich eine Richtung ab, die sich auf tendenzielle Probleme der Anthropomorphisierung von Robotern konzentriert. So betreffen grundsätzliche ethische Fragen, die sich im Zusammenhang der Robotik stellen, die Grenzziehung zwischen Mensch und Maschine. Sie ergeben sich aus dem Umstand, dass Roboter in bestimmten Hinsichten nicht klar von Bereichen des Menschlichen unterschieden werden. Vielmehr kann hier von einer „Ent-

grenzung“ gesprochen werden, die weniger durch die tatsächlichen technologischen Entwicklungen, sondern von eingespielten zwischenmenschlichen Interaktionen und auch – nicht zuletzt – durch Zukunftsvisionen in Fiktion, Populärwissenschaft und utopischen Technologiediskursen geprägt ist (Kehl & Coenen, 2016). Eigenschaften von Robotern, an denen entsprechende ethische Überschreitungen anschaulich werden, sind typischerweise diejenigen der Autonomie (im Sinne einer autonomen Handlungs- und Entscheidungsfähigkeit) und der anthropomorphen bzw. zoomorphen Gestaltung. Doch auch unabhängig von diesen Eigenschaften, die auf die jeweilige Gestaltung und Programmierung der Roboter zurückgehen, beinhaltet schon die Idee einer Mensch-Roboter-Interaktion eine Grenzüberschreitung. Denn im Gegensatz zur allgemeineren Mensch-Maschine-Interaktion, die sich an Verhältnisse des komplexen Werkzeuggebrauchs im Sinne von Steuerung oder Bedienung anlehnt, schwingt im Begriff der Interaktion die Bedeutung einer *Handlungsbeziehung* mit. Im Unterschied zu anderen technischen Geräten wird mit Robotern die Vorstellung verbunden, dass sie nicht nur als Werkzeuge verwendet und gesteuert werden, sondern dass sich eine echte kooperative oder kollaborative Zusammenarbeit im Team ergeben kann.

Ein anschauliches Beispiel für eine derartige Grenzüberschreitung kann am Begriff der Autonomie entwickelt werden. So schlagen Christaller und Kollegen (2001) eine Differenzierung von verschiedenen Konzepten der Autonomie vor, an der eine ethische Kluft zwischen Mensch und Maschine deutlich werden soll: Zwar sprechen wir sowohl in Bezug auf Roboter als auch in Bezug auf Menschen von ‚Autonomie‘, meinen damit aber jeweils etwas Anderes. Es gilt zu unterscheiden:

1. Autonomie erster Stufe (*technische Autonomie*): „Die Autonomie erster Stufe liegt in Fällen komplexer Automation mit technisch induzierten Freiheitsgraden vor. Die Eigenschaft der Autonomie bezieht sich dabei auf die Eigenschaft einer Maschine, in bestimmten Bewegungsräumen Steuerungen und Aktionen auszuführen.“ (Christaller et al., 2001, S. 126).
2. Autonomie zweiter Stufe (*personale Autonomie*): „Die eigentliche Autonomie bezeichnet eine Fähigkeit von Personen, spontan Einstellungen einzunehmen und Handlungen auszuführen, die prinzipiell nicht vorhersagbar sind. Personale Autonomie vollzieht sich in der Form von Handlungen im Raum der Gründe. Diese müssen nicht moralisch oder im engeren Sinn vernünftig bestimmt sein. Ein typischer Fall personaler Autonomie sind Lebenspläne im Sinne von Wünschen und Interessen zweiter Stufe.“ (Christaller et al., 2001, S. 126).

An dieser Aufschlüsselung des Autonomie-Begriffs soll deutlich werden, dass nur die Autonomie erster Stufe für die Robotik relevant ist. Die dargestellte Mehrdeutigkeit des Autonomiebegriffs könnte aber zu falschen Vorstellungen und Erwartungen führen, indem die (technische) Autonomie von Robotern fälschlicherweise mit Vorstellungen von personaler Autonomie vermischt wird. Personale Autonomie wird aber ethisch völlig anders bewertet als technische Autonomie: Die freie Wahl von Lebensplänen „im Sinne von Wünschen und Interessen zweiter Stufe“ ist ein Wert, dessen Verletzung (durch Zwang oder Manipulation) ethisch verwerflich ist. Dagegen sind Eingriffe in die Autonomie des Roboters offensichtlich ethisch unproblematisch.

Das Beispiel veranschaulicht, wie die verwendete Begrifflichkeit unter bestimmten Umständen zu einer ‚vermenschlichten‘ Rede über Roboter beitragen könnte.³ Doch nicht nur die Begrifflichkeit, sondern auch beispielsweise humanoide oder zoomorphe Erscheinungsformen von Robotern können in dieser Hinsicht ethisch relevant werden. Sie können etwa zu manipulativen Zwecken eingesetzt werden, beispielsweise indem gefühlsmäßige Anreize zur Preisgabe privater Daten geschaffen werden oder indem bestimmte Verhaltensweisen provoziert werden, die in Bezug auf Menschen nicht wünschenswert wären (Darling, 2015).

Eine allgemeine Lösung für derartige Probleme besteht darin, mit einer ethisch reflektierten ‚Rahmung‘ (*framing*) die Assoziationen und Verhaltensweisen interagierender Personen zu regulieren. Ein solches framing kann schon durch einfache Vorgänge wie beispielsweise Metaphern, Namensgebung oder fiktive Hintergrundgeschichten erfolgen. Daraus ergeben sich bestimmte Effekte, die zu einem ethisch wünschenswerten Umgang mit Robotern beitragen können (Darling, 2015). Auf diese Weise können Mensch-Roboter-Interaktionen also in ethischer Hinsicht gestaltet werden. Auch kann ein gezieltes framing das Bewusstsein für manipulative Mechanismen fördern, wodurch Effekte der Roboter-Gestaltung auf menschliches Verhalten transparent werden. Oder es kann eine unnötige Reproduktion von stereotypischen Geschlechterbildern im Roboter-Design vermieden werden. Ethisch orientiertes framing erscheint folglich als geeignete Methode zur Untersuchung und Gestaltung verschiedener Effekte der Mensch-Roboter-Interaktion (Vgl. Kapitel 4.6.4).

2.4 Zwischenfazit

Die Erforschung der ethischen und sozialen Implikationen der Robotik und insbesondere der Mensch-Roboter-Interaktion steht in Bezug auf die Konsequenzen der Anthropomorphisierung von Robotern noch am Anfang. Während Fragen nach der ethisch-moralischen Qualität autonomer Entscheidungen und nach der Verantwortungsfähigkeit von Robotern bereits intensiv diskutiert und in Anforderungen an rechtliche Bestimmungen übersetzt werden, sind die Technikfolgen der sozialen Dimension der MRI (beispielsweise in Arbeitsverhältnissen) noch nicht klar abzusehen. Während Beziehungen zu Artefakten oder auch fiktiven Figuren in gewissen Grenzen für ethisch unproblematisch gehalten werden (Carpenter, 2016), unterscheiden sich Roboter durch ihre Fähigkeit zu realen und verkörperten Handlungen und Entscheidungen grundsätzlich von diesen bekannten Konstellationen. Unabhängig von unrealistischen Spekulationen über eine ‚starke KI‘ in Robotern stellen sich Fragen nach den graduellen und schleichenden Veränderungen der Lebens- und Arbeitswelt durch den umfassenden Einsatz von konkreten Mensch-Roboter-Interaktionen. In diesem Bereich sind noch keine radikalen Beispielszenarien in Sicht, an denen die ethischen Probleme der technologischen Entwicklung sofort anschaulich werden (wie es beispielsweise das diskutierte Trolley-Dilemma für die Diskussion ethisch-moralischer Entscheidungen leistet). Daher wird im folgenden Kapitel 3 der Versuch unternommen, Tendenzen für kritische ethisch-soziale Entwicklungen der MRI im Arbeitskontext zu identifizieren. Dabei handelt es sich um eine argumentative Skizze von Möglichkeitsspielräumen, die sich im Ausgang von aktuell verwendeten Begrifflichkeiten der MRI eröffnen. An die Abschätzung möglicher ethisch-sozialer Aspekte

³ Für den rechtlichen Bereich wurde dieser Effekt als „The Android Fallacy“ bezeichnet (Richards & Smart, 2016).

werden anschließend in Kapitel 4 empirische Studien hinsichtlich bestimmter ethisch-sozialer Aspekte in Abhängigkeit von der Taxonomie nach Onnasch und Kollegen (Onnasch et al., 2016) ausgewertet.

3 Kritische Aspekte der MRI aus ethisch-sozialer Perspektive

Mensch-Roboter-Interaktionen erscheinen in vielerlei Hinsicht als vorteilhaftes Szenario für die Zukunft der Arbeit. Im Vordergrund der Vision steht die MRI als Ausweg aus der weiteren Vernichtung von Arbeitsplätzen durch die fortschreitende industrielle Entwicklung. Vom „Kollegen Roboter“ erwartet man keine Ersetzung des Menschen, sondern vielmehr hilfreiche Mitarbeit.⁴ Dieser Aspekt wird als soziale Implikation besonders in der öffentlichen Debatte hervorgehoben. Auf der anderen Seite können einige prinzipielle Produktivitäts- und Effizienzsteigerungen durch den Einsatz von Robotern genutzt werden. Denn in der MRI können Tätigkeiten durch Roboter ermöglicht und unterstützt werden, die dem Roboter auf dem gegenwärtigen Stand der Entwicklung ohne menschliche Kooperation und Kollaboration nicht möglich wären. Umgekehrt gilt dasselbe natürlich für den Menschen. Mit der MRI werden also neue Arbeitsfelder für die (Teil-)Automatisierung erschlossen.

Der vergleichsweise ‚gemäßigte‘ Einsatz von Robotern in MRI-Szenarien wirft neben sozialpolitischen Fragen der Verteilungsgerechtigkeit allerdings auch Fragen nach dem ethisch-sozialen Status der neuen Arbeitsbeziehungen auf. Wie in vielen Rationalisierungsstrategien können auch hier Effizienz-Aspekte unter Umständen in Konflikt mit einer positiv bewerteten Arbeitsgestaltung geraten. Die Besonderheit der MRI besteht in dieser Hinsicht darin, dass Menschen Maschinen nicht mehr nur als Werkzeuge verwenden, sondern in bestimmte Arbeitsbeziehungen zu Robotern eintreten, wodurch eingespielte Praktiken und Werte der Teamarbeit in einem anderen Licht erscheinen.

In diesem Sinne orientieren sich die folgenden ethisch-sozialen Beurteilungen an den Werten der Persönlichkeitsentfaltung und der Gesellschaftsqualität, wie sie in VDI 3780 definiert werden: „Persönlichkeitsentfaltung bedeutet, dass der Mensch seine Anlagen, Fähigkeiten und Neigungen im Wechselspiel mit seiner Umgebung so weit wie möglich verwirklicht; als soziales Wesen kann er dies nur im Zusammenleben und Zusammenwirken mit anderen Menschen.“ – „Die Beschaffenheit der zwischenmenschlichen Beziehungen sowie der überpersönlichen Verhältnisse und Einrichtungen, die aus diesem Zusammenwirken von Individuen und Gruppen entstehen, bezeichnet man als Gesellschaftsqualität.“ (Verein Deutscher Ingenieure 3780, S.20). In beiden Definitionen wird die besondere Bedeutung sozialer Beziehungen für eine Bewertung von Technologien betont. Folglich ergeben sich besondere und neuartige Fragen, wenn gezielt nicht mehr Personen, sondern Roboter als Interaktionspartner des Menschen in Betracht gezogen werden. So ist die Persönlichkeitsentfaltung un-

⁴ Die Ersetzung von Menschen durch Roboter wird gelegentlich als radikale Konsequenz aus der Anwendung von Robotern zur Ersetzung einzelner menschlichen Tätigkeiten gesehen. Diese Assoziation ist inspiriert von Science-Fiction-Szenarien, in denen Roboter die Menschheit als Ganze gewissermaßen ‚überwinden‘. Als durchaus positiv gemeinte Vision werden in populär-wissenschaftlichen Publikationen Ideen in dieser Richtung entwickelt; sie spielen für die tatsächliche aktuelle Forschung allerdings keine tragende Rolle. Andererseits ist aber die Ersetzung von Arbeitskräften durch Roboter ein bedeutsames und seit dem Einsatz von Industrierobotern viel diskutiertes Thema. So werden in der Technikfolgenforschung und in Beiträgen zu sozialen Implikationen der Robotik intensiv die sich abzeichnenden Veränderungen der menschlichen Arbeit thematisiert. Die damit verbundenen Problematikationen sind nicht neu und werden in Bezug auf die Robotik bereits seit den 70er Jahren diskutiert; vgl. Christaller et al. (2001), sowie aktuell Campa (2015).

mittelbar betroffen von der systembedingten Arbeitsqualität und von den Entscheidungsspielräumen des Menschen in der MRI, während die Gesellschaftsqualität durch die spezifisch ‚künstliche‘ Beziehung zwischen Mensch und Roboter beeinflusst werden kann.

Im Folgenden werden diese Konsequenzen genauer bestimmt und vor dem Hintergrund der MRI-Taxonomie nach Onnasch und Kollegen (2016) auf bestimmte Einflussgrößen bezogen, die einen Gestaltungsspielraum nach ethisch-sozialen Kriterien ermöglichen. Entsprechende MRI-Szenarien können damit in Relation gesetzt werden zu bestimmten Kategorien der Taxonomie wie Interaktionsrolle, Autonomie und Morphologie (vgl. Kapitel 4.3). Ethisch-soziale Auswirkungen werden schließlich durch Interpretationen geeigneter Studien bestimmt, deren Ergebnisse sich auf die theoretisch bestimmten Einflussgrößen beziehen lassen (vgl. Kapitel 4.4-4.6).

Zur Klärung des Ansatzes werden in einem ersten Schritt die ethischen und sozialen Implikationen der MRI in zwei Dimensionen eingeteilt. Die erste Kategorie umfasst diejenigen Auswirkungen auf den ethischen und sozialen Wert der menschlichen Arbeit, die durch die jeweiligen Rahmenbedingungen der Mensch-Roboter-Interaktion bestimmt werden. Aspekte, die in diese Dimension fallen, betreffen beispielsweise die Selbstbestimmung des menschlichen Arbeiters, den Ausdruck der Wertschätzung seiner Arbeitskraft und damit das sinnstiftende Potential der Arbeit. Probleme entstehen dann, wenn eine konkrete MRI die Persönlichkeit des menschlichen Mitarbeiters einschränkt oder nicht angemessen fördert.

Die zweite Dimension umfasst Aspekte der MRI, die bestimmte Einstellungen des Menschen in Bezug auf den Roboter beeinflussen. Ausgangspunkt dafür ist im Bereich der industriellen Produktion insbesondere die Zuschreibung intentionaler Zustände. Darüber hinaus beeinflussen aber auch morphologische Gestaltungen und Metaphern den Umgang mit Robotern. Im Hintergrund dieser Entwicklung steht ein ethischer Grundkonflikt von Mensch-Roboter-Interaktionen: MRI setzt Zuschreibungen typisch menschlicher Eigenschaften zu Robotern voraus wie z. B. Absichten, Entscheidungen und Handlungen, während bei rationaler Betrachtung Roboter Maschinen bzw. komplexe Werkzeuge sind, die nichts Menschliches an sich haben. Aspekte, die in diese Dimension fallen, betreffen beispielsweise das Verantwortungsbewusstsein des menschlichen Arbeiters, suggestive Spielräume für negative oder positive Verhaltensweisen und affektive Bindungen zu Robotern, die die Interaktion prägen. Probleme entstehen dann, wenn die Grenze zwischen Mensch und Roboter durch den Kontext Interaktion auf eine Weise verschwimmt, die Auswirkungen auf menschliche Handlungen im Besonderen und die Qualität des Zusammenlebens im Allgemeinen hat.

Werte der Technikfolgenabschätzung, die aus den folgenden Untersuchungen ausgegrenzt werden, sind insbesondere Sicherheit, Gesundheit und Umweltqualität (vgl. Verein Deutscher Ingenieure, 3780). Zwar sind angesichts der Einführung neuartiger Technologien in diesen Bereichen unter Umständen auch ethische Abwägungen erforderlich, beispielsweise wenn neue Risiken in Kauf genommen werden. Für den Rahmen des vorliegenden Berichts ergeben sich allerdings keine hinreichend spezifischen Fragestellungen. Während Sicherheitsanforderungen an kollaborative Roboter unter anderem in der ISO/TS 15066:2016 festgelegt sind, sind bisher keine spezifischen Auswirkungen von MRI-Szenarien auf Gesundheit und Umweltqualität in der

industriellen Produktion erkennbar, die sich wesentlich von allgemeineren Auswirkungen hochtechnologischer Automatisierung unterscheiden.

Besondere Probleme ergeben sich für den Bereich des Datenschutzes. Denn als informationstechnische Systeme registrieren und verarbeiten Roboter große Mengen an Daten. Fragen nach dem Umgang mit den gesammelten Daten ergeben sich besonders im Rahmen von Interaktionen, in denen unter Umständen aus nächster Nähe Informationen über den menschlichen Interaktionspartner registriert, gespeichert und verarbeitet werden müssen. Es eröffnen sich hier neue Überwachungs- und Kontrollmöglichkeiten, die im Rahmen einer Abwägung verschiedener Grundwerte zu beurteilen sind. Informationelle Selbstbestimmung, Schutz der Privatsphäre und verantwortungsvolle Verarbeitung der gespeicherten Daten stehen Sicherheitsmaßnahmen und Optimierungszielen gegenüber. Die mit diesen Aspekten verbundenen Fragen und Probleme werden im Folgenden nicht näher untersucht, weil sie ein weites Themenfeld berühren, in dem Mensch-Roboter-Interaktionen keine grundsätzlich neuen Szenarien eröffnen.

3.1 Ethisch-soziale Qualität der Arbeit in der MRI

Die Idee einer explizit sozialen Dimension der Arbeitsverhältnisse in der automatisierten industriellen Produktion umfasst neben der Gestaltung von Aspekten wie Sicherheit, Kontrolle und Ergonomie im MRI-Bereich spezifisch auch Qualifikationsbedürfnisse, technische und organisationale Kompetenzen. Diese Auffassung des sozialen Bereichs der Arbeitsgestaltung betrifft in erster Linie technische Verbesserungen und Sicherheitsanforderungen (Decker, 2013; Moniz & Krings, 2016). Darüber hinaus kann aber auch die ethisch-soziale Qualität der Tätigkeiten im Arbeitsprozess thematisiert werden. Aus der Perspektive des Arbeiters, der mit Maschinen und Robotern interagiert, drückt sich diese ethisch-soziale Qualität der Arbeit in Begriffen wie Sinnstiftung, Wertschätzung oder Identifikation mit der Arbeit aus. Negative Bewertungen werden hingegen mit Begriffen wie Instrumentalisierung, Entfremdung oder Monotonie bezeichnet. Entsprechende Werte können mit subjektiven Kriterien wie Motivation, Engagement und Zufriedenheit in Verbindung gebracht werden.

Zunächst kann jede Arbeitsteilung zu Ungleichgewichten oder zu unangemessenen Belastungen führen. Im Zuge von Rationalisierungsmaßnahmen zur effizienteren Arbeitsteilung kann die ethisch-soziale Qualität der Arbeit beispielsweise durch einen *down-skill*-Effekt leiden, indem die Nachfrage nach komplexeren menschlichen Tätigkeiten sinkt, während einfachere Tätigkeiten vermehrt gefordert sind (Decker, 2013). Dies kann zur Folge haben, dass positiv bewertete Fähigkeiten wie Problemlösungskompetenzen oder Kreativität weniger stark eingesetzt und ausgebildet werden. Daraus ergibt sich beispielsweise „die Gefahr einer Instrumentalisierung der menschlichen Arbeiter im Fertigungsprozess“: „Wenn sich in einem Fertigungsprozess die Kombinationen aus menschlichen und robotischen Tätigkeiten so darstellen, dass der menschliche Arbeiter nur noch die nicht rentabel robotisierbaren Übergangsarbeiten übertragen bekommt, kann das durchaus einer nicht akzeptablen Instrumentalisierung gleichkommen.“ (Decker, 2013, S. 355). Der sinnstiftende Ausdruck der eigenen Arbeit wird eingeschränkt, Monotonie und Entfremdung sind die Folge. Aus der hier maßgeblichen Perspektive des menschlichen Mitarbeiters müssen daher bestimmte Bedingungen erfüllt sein, um eine ethisch-soziale Verbesse-

rung der Gesamtsituation zu erreichen. Ein Ziel besteht folglich darin, beanspruchungsoptimale Tätigkeitsspielräume für den menschlichen Mitarbeiter zu eröffnen. MR-Kollaborationen erscheinen zur Erreichung dieses Ziels hilfreich, wenn der Roboter monotone und für den Menschen unergonomische Aufgaben übernimmt. Damit wird nicht nur die Effizienz des Arbeitsprozesses gesteigert, sondern dem menschlichen Interaktionspartner auch unangemessene Arbeit erspart. Menschlichen Mitarbeitern werden dann im Idealfall Tätigkeiten zugeordnet, die ihren Fähigkeiten und Ansprüchen optimal entsprechen. Damit hat die MRI das Potential, auf dieser Ebene zu einer verbesserten ethisch-sozialen Arbeitsqualität beizutragen.

In Bezug auf Automatisierungsszenarien in der Produktion argumentieren beispielsweise Applin & Fischer (2015) aus ethisch-sozialer Perspektive gegen ein Ersetzungsszenario und für ein kooperatives Arbeitsverhältnis zwischen Menschen und Robotern. Demnach gelingen Arbeitsprozesse häufig nicht deshalb, weil die Arbeiter genau ihren Anweisungen folgen und entsprechend ‚programmiert‘ werden, sondern weil sie die Arbeitsabläufe entgegen den Anweisungen flexibel und eigenständig modifizieren. Da entsprechende Modifikationen nicht unbedingt unmittelbar der Effizienzsteigerung dienen, sondern vielmehr als Ausdruck des eigenen Handelns die Qualität der Arbeitserfahrung verbessern, werden sie oft verdeckt, so dass das Produkt schließlich als Ergebnis einer genauen Befolgung von Anweisungen erscheint (Applin & Fischer, 2015). Erst infolge einer technologischen Ersetzung menschlicher Arbeitskraft tritt der Verlust nicht-konformer Handlungen als Mangel auf, zumal dann keine selbständigen Anpassungen mehr vorgenommen werden. Die Lösung dieses Problems sehen Applin & Fischer darin, kooperative und kollaborative Szenarien in der MRI zu fördern. Das Arbeitsverhältnis zwischen Menschen und Robotern kann in diesem Kontext so gestaltet werden, dass die menschlichen Bedürfnisse zum Ausdruck der eigenen Arbeit in Form z. B. von spontanen Problemlösungen, Freiheit und Kreativität gewährleistet und gefördert werden. Diese Qualitäten tragen zu einem ethisch wünschenswerten Arbeitsumfeld bei. Eine entsprechende Strategie zur Gestaltung des Arbeitsprozesses ist also aus ethischen Gründen wünschenswert – im Gegensatz zur vollständigen Ersetzung der Arbeitskraft durch Roboter.

In diesem Sinne kann die Entwicklung von MRI-Szenarien als positive technologische Gestaltung der Arbeitswelt gesehen werden, die nicht in erster Linie neue ethisch-soziale Probleme aufwirft, sondern die Chancen für eine Verbesserung der ethisch-sozialen Arbeitsqualität in sich birgt. Es muss in dieser Hinsicht nicht alles möglichst so bleiben wie es ist, sondern die ethisch-soziale Qualität der Arbeit kann durch den Einsatz von Robotern gesteigert werden. Die MRI eröffnet die Möglichkeit, den Status Quo zu überwinden und positive Visionen für das Arbeitsleben der Zukunft zu entwerfen.

Beurteilung der ethisch-sozialen Verbesserung der Arbeitsqualität durch MRI

In Argumentationen gegen die ‚Ersetzung des Menschen‘ in der Produktion, wie sie von Applin & Fischer (2015) formuliert werden, erscheint die MRI als technologische Lösung von Problemen mit ethisch-sozialer Dimension. Zwar setzt ein sinnvoller Einsatz der MRI in der industriellen Produktion grundsätzlich eine Effizienzsteigerung gegenüber anderen Modellen der Arbeitsgestaltung voraus. Es ist aber davon auszugehen, dass auch in MRI-Szenarien grundsätzlich eine Abwägung zwischen Effizienzgesichtspunkten und der ethisch-sozialen Arbeitsqualität stattfinden muss. Denn die Vision einer Verbesserung der Arbeitsqualität durch den Einsatz kollabora-

tiver Roboter ist nicht alleine vom Einsatz einer besonderen Technologie, sondern vielmehr von der konkreten Gestaltung der Interaktion und von ihrem organisatorischen Einsatz abhängig. Wie bei technologischen Entwicklungen im Allgemeinen gilt auch für die MRI, dass grundsätzlich nur dann eine tatsächliche Verbesserung erfolgt, wenn die Interaktionen auch gezielt in Hinblick auf Werte wie Persönlichkeitsentfaltung und Gesellschaftsqualität eingerichtet und gestaltet werden. Die Entwicklung von MRI-Szenarien kann hier Möglichkeiten eröffnen, die unter Einbeziehung entsprechender Kriterien genutzt werden müssen. In dieser Hinsicht verhält es sich bei kollaborativen Robotern nicht anders als bei anderen technologischen ‚Verbesserungen‘ der ethisch-sozialen Situation: Nicht die Technologie alleine, sondern erst der richtige Umgang mit ihr schafft reale Fortschritte im Hinblick auf ethisch-soziale Werte. Dabei gilt: Fortschritte dieser Art können sich zulasten der prinzipiell realisierbaren Gesamteffizienz des Produktionsablaufs auswirken.

Bei der Untersuchung der MRI als Vision für die Zukunft der Arbeit erscheint es allerdings irreführend, sich an einer Auspielung von MRI-Szenarien gegen die drohende Vollautomatisierung (d.h. Abschaffung möglichst vieler Arbeitsplätze) zu orientieren. Denn um das ethisch-soziale Potential der MRI genauer abschätzen zu können, müssen verschiedene MRI-Szenarien verglichen und hinsichtlich ihrer ethisch-sozialen Konsequenzen beurteilt werden. Diese Orientierung wird die Auswahl der Studien in Kapitel 4 anleiten. Es sind MRI-Szenarien denkbar, in denen zwar bestimmte Aspekte der Produktion verbessert werden, die ethisch-soziale Qualität der Arbeit für den interagierenden Menschen allerdings ein niedriges Niveau erhält. Denn prinzipiell ist davon auszugehen, dass erst ein informierter, reflektierter und verantwortungsbewusster Einsatz einer neuen Technologie auch auf ethisch-sozialer Ebene Verbesserungen (oder zumindest keine unvertretbaren Verschlechterungen) mit sich bringen kann.

3.1.1 Entscheidungsautorität: Führende Funktionen von Robotern

In kooperativen und kollaborativen Szenarien der MRI besteht die Möglichkeit, den Handlungsspielraum des Menschen unmittelbar durch den Roboter zu determinieren. In diesem Zusammenhang kann davon gesprochen werden, dass der Roboter gegenüber dem Menschen eine ‚führende‘ Funktion im ‚hybriden Team‘ einnimmt (Krüger, Lien, & Verl, 2009).

Im Vergleich zu bisherigen Formen der technischen Strukturierung industrieller Arbeit, wie sie beispielsweise durch die Vorgabe einer Taktzeit geprägt sind, werden führende Funktionen von Robotern spezifisch relevant, wenn sie eine besondere interaktive Qualität haben. Dies ist beispielsweise dann der Fall, wenn der Roboter einen hohen Autonomiegrad aufweist, so dass die autonomen Entscheidungen des Roboters die Tätigkeit des Menschen bestimmen. Beispielsweise kann in einem Arbeitsablauf, in dem ein Roboter menschlichen Interaktionspartnern nach einer Lern- und Programmierphase situationsabhängig bestimmte Arbeitsschritte vorschreibt, durch die Einbeziehung von Effizienzberechnungen eine möglichst optimale Arbeitsteilung realisiert werden. Entsprechende Berechnungen können im Rahmen einer informationstechnischen Verarbeitung seitens des Roboters schneller und genauer ausgeführt werden, als es einem menschlichen Interaktionspartner möglich wäre. Darüber hinaus erscheinen die Anweisungen eines Roboters u. U. ‚selbständiger‘ und geplanter als vergleichbare Arbeitsführungen von automatischen Systemen

(z. B. Fließbänder). Man kann dieses Szenario daher auch so deuten, dass der Roboter die Tätigkeit eines Koordinators ersetzt, der sich bei der selbständigen Planung am Maßstab computergestützter Effizienzberechnungen orientiert. Die führende Funktion des Roboters erscheint in dieser Hinsicht zunächst unproblematisch: Die Anweisungen des Roboters sind förderlich für den Arbeitsprozess und die Erreichung der Ziele des Mensch-Roboter-Teams.

An diese Charakterisierung knüpft die Frage an, ob in diesem Zusammenhang Begriffe wie „Führungsrolle“ oder „Team-Leitung“ in Bezug auf Roboter angemessen sind. Zunächst: Beschreibt diese Begrifflichkeit die Interaktionen zwischen Menschen und Robotern adäquat? Wäre die Taxonomie der Mensch-Roboter-Interaktionen um eine ‚führende‘ Interaktionsrolle des Roboters zu erweitern?

Wendungen wie „Führung“ von Interaktionspartnern im Produktionsprozess oder „Leitung“ im Team implizieren im weitesten Sinne eine Einwirkung auf komplexes menschliches Verhalten (Wagner, Rex, & Eicher, 2003). In der üblichen Anwendung auf menschliche Interaktionen beruhen Führungstätigkeiten immer auch auf sozialen Kompetenzen. So werden nicht nur Entscheidungen getroffen, sondern es ergibt sich beispielsweise auch ein Spielraum dafür, wie diese Entscheidungen kommuniziert und in den umfassenderen Arbeitskontext integriert werden. Vergleichbare Kompetenzen sind in autonomen Robotern offenbar nicht implementiert, so dass die entsprechenden autonomen Entscheidungen nicht unmittelbar eine ‚Führungsrolle‘ implizieren. Die autonome Entscheidung des Roboters stellt sich eigentlich als maßgebliches Ergebnis einer Datenverarbeitung dar, die im Setting der MRI zur unmittelbaren Handlungsanleitung dienen kann. Ein starker Begriff wie der einer ‚Führungsrolle‘ erscheint in diesem Zusammenhang irreführend, denn die Befolgung von ‚Anweisungen‘ eines Roboters unterscheidet sich nicht grundsätzlich von der Orientierung an der Ausgabe technischer Instrumente, beispielsweise technischer Informationsverarbeitungen, Taktgebern u. ä.

Neben der Ausklammerung sozialer Kompetenzen wäre auch eine Trennung zwischen Führungsrolle und der damit einhergehenden Verantwortlichkeit problematisch. Wie in Abschnitt 2.4 dargestellt, können Roboter prinzipiell keine Verantwortung für autonome Entscheidungen tragen. Dieser Umstand wirkt bei der Zuschreibung von Leitungs- oder Führungsrollen besonders problematisch. Denn gerade in Szenarien, in denen die Steuerung der menschlichen Tätigkeit an autonome Roboter delegiert wird, liegt die Forderung nach verantwortungsvollen Entscheidungen nahe. Eine entsprechende verantwortliche ‚Autorität‘ eines führenden Roboters kann bestenfalls eine Vermittelte sein, sofern im Hintergrund eben ein menschlicher Vorgesetzter den Roboter kontrolliert. Die Bereitschaft, eine derartige mittelbare Verantwortung zu übernehmen, dürfte aber mit zunehmender Autonomie des Roboters abnehmen.

Aus diesen Gründen sind vergleichsweise starke Begriffe wie der der Führungsrolle oder auch der Teamleitung unpassend, wenn sie auf Roboter in der MRI angewandt werden. Problematisch wird es, wenn durch die Wortwahl eine ‚Autorität‘ des Roboters suggeriert wird. Angemessen erscheint dagegen die Rede von einer führenden Funktion des Roboters, wenn damit eine Strukturierung der Arbeitsprozesse durch Automatisierung verbunden wird.

Doch auch unabhängig von den verwendeten Begriffen kann im Rahmen der MRI auch die *Wahrnehmung* einer Entscheidungsautorität bzw. einer (vermeintlichen) Führungsrolle des Roboters durch bestimmte Gestaltungsmerkmale gefördert werden. So ist beispielsweise eine subtile, aber dennoch gezielte Einflussnahme auf den Menschen möglich, so dass er geneigt ist, bestimmte Entscheidungen und Vorgaben des Roboters zu akzeptieren, obwohl sie eigentlich nicht im eigenen Interesse liegen. Die unwillkürliche Anerkennung einer ‚Führungsautorität‘ des Roboters verleitet so zu einer Interaktion, die den Menschen in gewissem Maße unbewusst manipuliert. In diesem Zusammenhang gilt es, Klarheit zu schaffen nicht nur über die verwendete Begrifflichkeit, sondern auch über die Auswirkungen von Gestaltungsmerkmalen der MRI auf die (unbewusste) Wahrnehmung einer Autorität des Roboters (vgl. Kapitel 4.4-4.6).

Doch auch dann, wenn man die führende Funktion eines Roboters auf reine Optimierungsleistungen beschränkt (beispielsweise bei der Anleitung zu einfachen Tätigkeiten in einem arbeitsteiligen Prozess), also nicht mit sozialen Kompetenzen und Verantwortung im starken Sinne einer ‚Führungsrolle‘ verknüpft, ergeben sich Fragen nach der Angemessenheit derartiger Szenarien. Zwar erscheint eine führende Kooperation/Kollaboration von Robotern in der MRI durchaus gerechtfertigt, sofern sie in Analogie zur Orientierung an der Ausgabe von Datenverarbeitungen verstanden wird. Roboter sind in bestimmten Tätigkeitsbereichen leistungsfähiger als Menschen; im genannten Beispiel kann ein Roboter die optimale Arbeitsverteilung aufgrund seiner Rechenleistung besser bestimmen. Der menschliche Interaktionspartner kann die Überlegenheit des Roboters in Bezug auf diese Tätigkeit einsehen und folgt seiner ‚Entscheidung‘. Eine Orientierung an den Vorgaben des Roboters in der MRI ist unter dieser Voraussetzung unproblematisch. Probleme wirft die Einführung derartiger Interaktionsverhältnisse zwischen Mensch und Roboter allerdings dann auf, wenn der Roboter in Bezug auf eine Entscheidung für ‚besser‘ gehalten wird als der Mensch, die zugrundeliegenden Fähigkeiten allerdings a) nicht gleichwertig oder b) nicht vergleichbar sind:

- a) Fähigkeiten von Robotern und Menschen sind nicht gleichwertig, wenn die Komplexität der menschlichen Fähigkeit bei der Entwicklung des Roboters unterschätzt wird. Ein Grund dafür kann beispielsweise darin liegen, dass der Entwicklung des Roboters ein vereinfachendes Modell der nachzubildenden menschlichen Tätigkeit zugrunde liegt. Nimmt der Roboter dann eine führende Funktion in der MRI ein, ergibt sich daraus für den interagierenden Menschen eine Herabsetzung der eigenen Fähigkeit. Was in der technischen Realisierung funktional und effektiv ist, kann aus der Perspektive eines erfahrenen Arbeiters defizitär und unterkomplex erscheinen. Erschwerend kommt hinzu, dass der menschliche Interaktionspartner nicht unbedingt über die theoretischen Mittel verfügen dürfte, die Unterlegenheit von Roboterentscheidungen und -tätigkeiten zu begründen. Denn sie äußert sich nicht unbedingt im Ergebnis, sondern vielmehr im Arbeitsprozess selbst.
- b) Unvergleichbar sind Fähigkeiten hingegen dann, wenn mit einer entsprechenden Entscheidung auch ein Ausdruck ethisch-sozialer Werte verbunden ist. Entscheidungen und Tätigkeiten von Robotern entbehren einer authentischen sozialen Dimension, so dass derartige Einflüsse auf eine führende Funktion entfallen. Gerade im Bereich der Optimierung von Tätigkeitsabläufen liegt das

Beispiel einer Anpassung an individuelle und möglicherweise sogar persönliche Eigenschaften von Mitarbeitern nahe. Beispielsweise könnte ein menschlicher Vorgesetzter auf bestimmte Befindlichkeiten eines Mitarbeiters Rücksicht nehmen, etwa wenn dieser in persönlichen Schwierigkeiten steckt. Hier sind wiederum soziale Kompetenzen erforderlich, die einem Roboter in absehbarer Zeit nicht einprogrammiert werden können. Die Fähigkeit, über den ethisch-sozialen Wert solcher Entscheidung zu urteilen, liegt alleine beim Menschen. Der Maßstab wird in diesem Bereich vom Mensch gesetzt und tritt nicht als Ergebnis einer Berechnung auf; dies ist ein Ausdruck menschlicher Autonomie. Auf der Ebene ethisch-sozialer Werte können technisch-autonome Entscheidungen eines Roboters daher nicht als prinzipiell ‚besser‘ oder als gleichwertig anerkannt werden (vgl. dazu auch Kapitel 3.1).

In ähnlicher Weise sind auch Szenarien denkbar, in denen eine führende Funktion des Roboters in Widerspruch zu einer grundsätzlichen ethisch-sozialen Wertschätzung des menschlichen Interaktionspartners gerät. In diesem Falle ist von einer *Instrumentalisierung* des Menschen auszugehen. Beispielsweise könnte der Fall eintreten, dass die maximale Effizienz einer MRI, die durch den Roboter berechnet und vorgegeben wird, zu einer unangemessenen, monotonen und unterfordernden Tätigkeit des Menschen führt. Auch wenn der Mensch einsieht, dass der Roboter die ‚richtigen‘ (rationalen) Entscheidungen trifft, wirken sie sich in der Konsequenz nachteilig auf den Wert der Tätigkeit im Arbeitsprozess aus. Diese Konsequenz könnte u. U. dadurch vermieden werden, dass Parameter wie beispielsweise Motivation, Auslastung und Anspruch der menschlichen Tätigkeit in die Ermittlung des optimalen Tätigkeitsspielraums einbezogen werden. Dabei stellt sich allerdings die Frage, ob der ethisch-soziale Wert der Arbeit hinreichend operationalisiert und modelliert werden kann.

In den Konstellationen a) und b) setzt eine führende Funktion des Roboters den Wert menschlicher Fähigkeiten und Tätigkeiten herab. Die Forderung, einer technisch realisierten autonomen Entscheidung eines Roboters zu folgen, dürfte unter diesen Umständen als unmenschliche Unterordnung unter eine technisch-rationale Autorität beurteilt werden.

3.1.2 Kommunikation: Die soziale Dimension der Sprache

Die Metapher der ‚Führungsrolle‘ des Roboters im Team erscheint auch dadurch problematisch, da Anweisungen nicht im eigentlichen Sinne ‚besprochen‘ werden können. Einwände, Vorschläge, Nachfragen o. ä. sind nicht im geläufigen Sinne möglich und müssen in technische Anfragen ‚übersetzt‘ werden. Eine sprachliche Abstimmung des Arbeitsprozesses ist insofern unmöglich; die Kommunikation zwischen (führendem) Roboter und menschlichen Mitarbeitern vereinfacht sich daher in Kollaborationen und beschränkt sich auf das technisch Notwendige. Hier hat die Metapher des ‚Teams‘ allerdings eine Grenze: Die sprachliche Aushandlung und Reflexion gemeinsamer Mittel und Ziele, die für die Arbeit in einem (menschlichen) Team charakteristisch ist, entfällt.

Auch unabhängig von dem Szenario einer führenden Interaktion ist zu bedenken, dass außerplanmäßige bzw. unvorhergesehene Einwände oder Bedürfnisse des menschlichen Mitarbeiters nicht kommuniziert werden können, sofern sie nicht

sprachlich mitgeteilt werden können. Nicht-sprachliche Kommunikation dürfte in der MRI technisch auf Äußerungen beschränkt sein, die sich auf bestimmte Aspekte des Arbeitsprozesses beziehen. Die Kommunikation im Mensch-Roboter-Team erscheint in dieser Hinsicht stark reguliert und auf eng umrissene Zwecke eingeschränkt. Die Offenheit und Flexibilität der Umgangs- und Fachsprache kann auf dieser Ebene nicht eingebracht werden.

Ohne die Möglichkeit einer umgangs- oder fachsprachlichen Verständigung entfällt in der MRI schließlich auch die soziale Dimension der Arbeit. Das mag je nach konkretem Anwendungsfall positive oder negative Konsequenzen haben, es fördert ab einem bestimmten Ausmaß zumindest das Risiko von Mangelerscheinungen wie beispielsweise sozialer Isolation oder Entfremdung.

Die Idee, eine verbal-sprachliche Kommunikation mit Robotern zu simulieren (vgl. digitale Assistenten wie Siri oder Alexa), kann hingegen ebenfalls problematische Implikationen haben. Abgesehen davon, dass sie im Arbeitskontext im Vergleich zu anderen Kommunikationsweisen nicht effizient erscheint, sind bei der Ausgestaltung der Sprachfähigkeit besondere Entscheidungen zu treffen: Zum Beispiel sollte man sich bei Simulationen gesprochener Sprache darüber bewusst sein, dass die Wahl einer Stimme aufgrund der Assoziation mit geschlechtlichen Merkmalen auch Stereotype transportieren kann und möglicherweise Vorurteile unterstützt (Negativbeispiel: Roboter mit weiblicher Stimme, der nur einfache Assistenzaufgaben durchführt, während männliche Stimmen bei Robotern die Taktzeit vorgeben). Darüber hinaus ist auch zu klären, welche Beziehung durch die Kommunikation vermittelt werden soll: Kommuniziert der Roboter in Form von klaren Anweisungen oder soll durch die Kommunikation eine eher zwanglose und lockere Stimmung vermittelt werden? Dieser Einsatz verbaler Kommunikation führt im Übrigen zu einer tendenziellen Anthropomorphisierung, womit die im Folgenden Abschnitt 3.2 diskutierten Problemkontexte eröffnet werden.

In den meisten Anwendungsfällen liegt es allerdings nahe, anstatt einer verbalen Kommunikation bestimmte Arten non-verbaler Kommunikation mit dem Roboter zu realisieren, die über Gestik, „Mimik“ oder auch die Bewegung selbst vermittelt wird. Der Einsatz non-verbaler Kommunikation in der MRI erscheint im Gegensatz zu einer simulierten verbal-sprachlichen Kommunikation vorteilhaft und geboten. Auf diesem Wege kann vertrautes Vorwissen des Menschen für die Interaktion mit Robotern nutzbar gemacht werden. Ziel ist eine intuitivere und flüssige Kollaboration. Darüber hinaus können vermenschlichende Tendenzen in diesem Rahmen ebenfalls vermieden werden, da non-verbale Kommunikation nicht alleine aus der menschlichen Kommunikation geläufig ist, sondern auch in der Kommunikation mit Tieren, so dass hier keine eindeutigen menschlichen Assoziationen naheliegen.

3.2 Ethisch-soziale Implikationen humanoider und zoomorpher Projektionen

Menschen schreiben Robotern im Rahmen von Mensch-Roboter-Interaktionen Handlungen, Absichten und Ziele zu (Darling, 2015; Carpenter, 2016). Zuschreibungen (Attributionen) dieser Art sind zunächst nichts Außergewöhnliches. Sie lassen sich ganz allgemein im Umgang mit bestimmten Objekten beobachten, besonders deutlich im Umgang mit technischen Geräten. Hier finden wir neben emotionalen Reakti-

onen wie Empathie auch Zuschreibungen von elementaren intentionalen Zuständen bis hin zur Personifizierung. In der Wahrnehmung von Robotern werden entsprechende Projektionen (im Vergleich zu Maschinen im Allgemeinen) durch drei typische Eigenschaften unterstützt: 1. Durch die Körperlichkeit von Robotern (im Gegensatz zu Software wie z. B. dem digitalen Assistenten Siri von Apple), 2. durch selbständige bzw. autonome Bewegungen, die Absichten und Ziele implizieren, sowie 3. durch soziale Erscheinung oder entsprechendes Verhalten – etwa im Rahmen von anthropomorpher bzw. zoomorpher Gestaltung (Darling, 2012). Im Bereich der sozialen Robotik werden diese Zusammenhänge gezielt genutzt, indem verbreitete und spontane Neigungen zur Anthropomorphisierung von ‚künstlichen‘ Apparaten abgerufen werden. Ab einer gewissen Intensität der Interaktion kann dies eventuell auch eine Art „psychologischen Zwang“ zur Zuschreibung von Empfindungsfähigkeit bewirken (vgl. Birnbacher & Birnbacher (2016), die hier eine Vermutung Marvin Minskys aufgreifen).

Im Gegensatz zu sozialen Robotern orientiert sich die Gestaltung von Industrierobotern normalerweise an rein funktionalen Gesichtspunkten. Anthro- oder zoomorphe Gestaltungen dürften normalerweise kaum sinnvoll sein. Im industriellen Arbeitsumfeld erscheinen Projektionen der genannten Art daher zunächst paradox. Roboter haben in diesem Bereich bisher deutlichen Werkzeugcharakter, so dass anthropomorphisierende Zuschreibung unpassend wirken. In den hier üblichen Szenarien der MRI, in denen Roboter-Arme oder mobile Plattformen in direkter Nähe zu menschlichen Arbeitern agieren, entsteht vielmehr der Eindruck einer flüssigen Werkzeugverwendung.

Allerdings durchbricht die Begrifflichkeit der MRI mit Konzepten wie ‚Mensch-Roboter-Kollaboration‘, ‚hybriden Teams‘ oder ‚Robotern als Kollegen‘ die eindeutige Kategorisierung von Robotern als Werkzeugen. Konzepte wie diese beruhen nicht (oder zumindest nicht nur) auf psychologischen Tendenzen, die durch Gestaltung und Verhalten von Robotern abgerufen werden, sondern sie sind auch notwendig mit dem Konzept der Interaktion verbunden. Insbesondere in Kollaborationen müssen Menschen ihren künstlichen Interaktionspartnern bestimmte Ziele und Absichten (Intentionen) zuschreiben, um die gemeinsamen Handlungen überhaupt koordinieren zu können. Daher eröffnet der gezielte Einsatz von MRI im Arbeitskontext einen weiten Bereich sozialer und rationaler Handlungskontexte, in dem nicht nur Absichten, sondern auch Verantwortung, Werte und Handlungsziele ins Spiel kommen. Eine klare Trennung zwischen handlungsorientierten Projektionen und weitergehenden ethisch-sozialen Kontexten erscheint hier schwierig.

Auch im Bereich der industriellen Arbeit haben wir es folglich mit anthropomorphisierenden Zuschreibungen zu tun, die in kooperativen und kollaborativen Mensch-Roboter-Interaktionen ebenso wie im Umgang mit anthropomorphen Robotern gegeben sind. In dieser Hinsicht ist auch die in Kapitel 2 dargestellte Diskussion der moralisch-ethischen Entscheidungsdimension von Bedeutung, zumal gerade im Umgang mit autonom agierenden Robotern Projektionen von handlungs- und verhaltensrelevanten Eigenschaften erforderlich werden. Im Kontext der MRI sind folglich auch Aspekte zu berücksichtigen, die aus Einstellungen bezüglich einer (vermeintlich) ‚moralischen Maschine‘ entstehen. Denn ganz unabhängig davon, ob Roboter tatsächlich auf der Ebene ihrer Programmierung moralischen Regeln folgen, ethisch reflektieren oder eine Sensibilität für ethisch heikle Situationen besitzen, werden sie von Men-

schen in Interaktion häufig entsprechend aufgefasst und behandelt. Denn Menschen nehmen aufgrund der genannten Eigenschaften Haltungen gegenüber Robotern ein, aus denen heraus sie ganz unabhängig von der tatsächlichen Realisierung bestimmte Eigenschaften auf ihr künstliches Gegenüber projizieren. Roboter-Verantwortung, moralische Bewertungen und ethische Konflikte spielen daher auch dann eine Rolle, wenn sie aus einer rationalen Perspektive keine Grundlage haben.

3.2.1 Verantwortungsdiffusion

Eine problematische Konsequenz aus den genannten Projektionen kann darin bestehen, dass Verantwortungszuschreibungen erschwert werden. Autonomes Verhalten, anthropomorphe Gestaltung, aber auch kollaborative Strukturen der Interaktion können die transparente Zuordnung von Handlungsursachen einschränken. Eine Beurteilung von Verantwortungsfragen erscheint dann nicht nur für den menschlichen Interaktionspartner, sondern sogar aus der Außenperspektive schwierig. Derartige Situationen führen zu einer *Diffusion* der Verantwortung: Es wird unklar, wer für Ergebnisse der Interaktion zuständig, verantwortlich und auch haftbar ist.

Verantwortungsdiffusion ist ein bekanntes Phänomen in Team- und Gruppeninteraktionen. Im Vergleich zu entsprechenden Effekten in menschlichen Teams erhält die Verantwortungsdiffusion in der MRI allerdings eine neue Qualität: Denn einerseits wäre es sicherlich unverhältnismäßig, wenn der menschliche Interaktionsteilnehmer die Verantwortung für technische Fehler des Roboters übernehmen müsste. Andererseits kann der Roboter als technisches Gerät gar keine reale Verantwortung tragen. Stattdessen kommen die Hersteller und Programmierer des Roboters oder die Besitzer (wie bei Tieren) als Verantwortungsträger in Frage, die an der konkreten Interaktion allerdings nicht beteiligt sind. Durch diese Verlagerung der Verantwortung und durch die damit verbundene Abstraktion vom konkreten Handlungsgeschehen werden die bisher üblichen Zuordnungen verkompliziert, insbesondere in Interaktionen mit selbstlernenden Systemen.

Unklarheit über Verantwortungsfragen kann im Verlauf der Interaktion verschiedene Konsequenzen haben. Zunächst ist ein Vertrauensverlust zu erwarten. Es wird grundsätzlich mit der Möglichkeit gerechnet, dass sich der Roboter in der Interaktion unerwartet oder sogar fehlerhaft verhält, während offenbar zugleich kein interner Zusammenhang zwischen dem Roboterverhalten und der Verantwortungsdimension besteht. In dieser Konstellation erscheint beispielsweise eine Distanzierung des menschlichen Interaktionspartners von der Handlung denkbar. Wenn die eigene Zuständigkeit für Ergebnisse der Interaktion nicht richtig eingeschätzt werden kann, dann ist eine risikomindernde Regulierung der eigenen Leistung folgerichtig.

Die Entwicklung weiterer Verhaltenstendenzen im Handlungsspielraum der MRI liegt nahe. Denn indem neue Technologien neue Handlungsmöglichkeiten eröffnen, machen sie auch neue Praktiken und Verhaltensweisen erforderlich. In den Handlungsspielräumen, die durch die Mensch-Roboter-Interaktion eröffnet werden, bieten sich bestimmte Verhaltensweisen und Umgangsformen an, die aus der Perspektive des Interaktionspartners vernünftig und vorteilhaft erscheinen, tatsächlich aber den Gesamtnutzen schmälern. Das Bewusstsein darüber, dass Roboter keine Verantwortung haben, dürfte einen wesentlichen Aspekt für dysfunktionale Verhaltensweisen ausmachen. Wenn niemand unmittelbar für das Funktionieren des Roboters verant-

wortlich zu machen ist (bzw. die Verantwortlichen nicht anwesend sind), wird dies auch die Einstellung zur Interaktion beeinflussen.

Verfügt der Roboter über einen bestimmten Grad an Entscheidungs- und Handlungsautonomie, können sich in der Interaktion etwa menschliche Verhaltensweisen einstellen, in denen der menschliche Interaktionspartner beispielsweise bestimmte Schwächen der Programmierung des Roboters zum eigenen Vorteil ausnutzt, wodurch sich die Interaktion negativ auf Außenstehende auswirkt. In diesem Szenario erfüllt der Roboter zwar seine Aufgabe entsprechend seiner Programmierung, es wird aber eine Art ‚Schlupfloch‘ in der Interaktion gefunden, wodurch bestimmte Eigenschaften der Roboterfunktion zum Nutzen des menschlichen Interaktionspartners gewendet werden. Entsprechende Tendenzen wurden im Zusammenhang mit autonomen Fahrzeugen im Straßenverkehr untersucht. Neben Missbrauchsszenarien sind vorhersehbare Fehlanwendungen und einschleichende Complacency (Übervertrauen in den Roboter) bei der Gestaltung von konkreten MRI-Szenarien zu berücksichtigen.

3.2.2 Affektive Bindungen des Menschen an den Roboter

Die Überschreitung des Werkzeugcharakters von Robotern erreicht ihr höchstes Maß, wenn sich im Rahmen der Interaktion eine affektive Bindung des Menschen zum Roboter einstellt. Bereits die Begrifflichkeit des ‚hybriden Teams‘ oder des ‚Kollegen Roboters‘ legt entsprechende Beziehungen nahe. In der Konsequenz erfordert die Konzeption einer Interaktion zwischen Menschen und Robotern gewisse Handlungs- und Verhaltensweisen, die aus der Mensch-Mensch-Interaktion geläufig sind, wie z. B. die Zuschreibung von Handlungsabsichten und Zielen. Starke Einflüsse auf eine soziale Wahrnehmung von Robotern sind im Übrigen auf die morphologische Gestaltung, autonome Bewegung und simuliertes soziales Verhalten zurückzuführen (vgl. Abschnitt 3.2).

Für sich genommen sind soziale Projektionen auf Roboter ethisch nicht unbedingt problematisch (Carpenter, 2016). Schwierigkeiten ergeben sich allerdings dann, wenn bestimmte Grenzen überschritten werden und die affektive Bindung den funktionalen Einsatz von Robotern beeinträchtigt. So kann sich z. B. bei einer übertriebenen Wertschätzung des „Kollegen Roboters“ die Abwägung von Vor- und Nachteilen einer Handlung zugunsten des Roboters verschieben. Ein extremes Szenario könnte etwa darin bestehen, dass Menschen eine Hemmung entwickeln, den Roboter in einer Gefahrensituation abzuschalten oder gar zu zerstören. Hier kann ein unbewusstes Zögern relevant werden, selbst wenn es nur um Sekunden geht (Darling, Nandy, & Breazeal, 2015).

Darüber hinaus hat die soziale Einbettung von Robotern nicht nur affirmativen Charakter. Es kann auch Ablehnung von Robotern als ‚Kollegen‘ bzw. Teampartner auftreten, die auf bestimmte Einstellungen zur Technik im Allgemeinen oder zur Robotik im Speziellen zurückgeführt werden kann, wie beispielsweise Ängste, Überforderung und Unkenntnis. Hier könnte eine suggestiv soziale Gestaltung der MRI kontraproduktiv wirken, wenn sie dem Roboter aus Sicht des Menschen einen künstlichen oder unbeholfenen Charakter verleiht.

Festzuhalten ist schließlich, dass soziale Beziehungen zu Robotern in der MRI nicht als Ersatz für den Verlust der sozialen Qualität der Arbeit dienen sollten. Maßnahmen in dieser Richtung würden den Werkzeugcharakter des Roboters mit einer sozialen Dimension anreichern, die wohl kaum die soziale Qualität von menschlichen Interaktionen erreichen kann. Da die soziale Funktion der MRI in Arbeitskontexten nicht im Vordergrund steht, sollten sich sozialisierende und personifizierende Gestaltungsmaßnahmen generell auf einen funktionalen und entsprechend minimalen Einsatz beschränken.

4 Analyse von MRI-Studien in Bezug auf ethische und soziale Aspekte

Ziel der folgenden Analyse ist es, die vorab aufgezeigten kritischen ethisch-sozialen Aspekte der Mensch-Roboter-Interaktion auf konkrete Gestaltungsmerkmale der Zusammenarbeit von Menschen und Robotern zu beziehen. Somit können Aussagen darüber getroffen werden, welche Merkmale der Zusammenarbeit besonders relevant für das Entstehen ethisch-sozialer Probleme sind, aber auch, welche Gestaltungsaspekte der Interaktion aus dieser Perspektive vorteilhaft sind.

Um dieses Ziel zu erreichen, setzt das methodische Vorgehen top-down an, d.h., die Studienauswahl erfolgt anhand der in Kapitel 3 diskutierten übergeordneten ethischen Prinzipien, Probleme und Aspekte. Es werden experimentelle Studien aus wissenschaftlichen Fachzeitschriften (z. B. *International Journal of Social Robotics, Autonomous Robots* oder *Human-Computer Interaction*) und Konferenzbeiträge (z. B. *Proceedings of the IEEE Symposia on Robot and Human Interactive Community*) analysiert, die den Einfluss verschiedener Gestaltungsaspekte der MRI im weitesten Sinne entweder auf die ethisch-soziale Qualität der Arbeit (s. Kapitel 3.1) oder auf das Projektionspotential ethischer Eigenschaften des Roboters untersucht haben (s. Kapitel 3.2). Um in die Analyse einzugehen, müssen Studien mindestens ein Merkmal der MRI systematisch variieren (z. B. die Autonomie oder Morphologie des Roboters).

Die in den jeweiligen Studien berichteten Wirkgrößen (unabhängige Variablen) werden anhand der MRI-Taxonomie von Onnasch und Kollegen (2016) systematisiert. Gleichzeitig werden die neu gewonnenen Erkenntnisse auch genutzt, um die MRI-Taxonomie weiter zu verfeinern. Der Erkenntnis- und Analysefluss erfolgt an dieser Stelle also bidirektional.

Die in diesem Kapitel berichteten Arbeiten stellen die Grundlage der Modellentwicklung (Kapitel 5) dar und werden im Folgenden detailliert erläutert.

4.1 Spezifizierung ethischer Aspekte

In Kapitel 3 wurden als ethisch relevante Dimensionen die ethisch-soziale Qualität der Arbeit sowie die Projektion ethischer Eigenschaften auf Roboter diskutiert. Während sich die erste Dimension mit Auswirkungen von MRI auf die generelle Gestaltung des Arbeitsumfelds beschäftigt, also mit Rahmenbedingungen der Interaktion zwischen Mensch und Roboter, fokussieren assoziierte Themen ethischer Eigenschaftsprojektionen auf Roboter-Aspekte der direkten Interaktion, die durch Einstellungen und Verhaltensweisen des interagierenden Menschen beeinflusst werden. Während also die in Kapitel 3.1 diskutierten Aspekte Vorgaben der Arbeitsgestaltung darstellen und Arbeitsplatzinhabern wenig Spielraum lassen, handelt es sich bei den in Kapitel 3.2 diskutierten Merkmalen um durch den Roboter und die Interaktion entstehende Angebote und Möglichkeiten (sowohl positiv als auch negativ), die der Mensch annehmen kann oder nicht und deren Ausgestaltung individuell unterschiedlich ausfallen kann.

Bezüglich der *ethisch-sozialen Qualität der Arbeit* werden für die Analyse die folgenden zwei Aspekte aus Kapitel 3 aufgegriffen:

1. Entscheidungsautorität und Instrumentalisierung (vgl. Kap. 3.1.1)

Eine Orientierung am Grundwert der menschlichen Autonomie (Selbstbestimmung) erkennt den Menschen als Maßstab für die Gestaltung der MRI an. Neuartige Probleme können in der MRI allerdings entstehen, wenn Roboter in der Interaktion eine führende Funktion ausüben. Entsprechende Funktionen des Roboters können in bestimmten Arbeitskontexten mit der ethisch-sozialen Qualität der Arbeit kollidieren, sofern dem menschlichen Arbeiter Tätigkeiten übertragen werden, die seinen Fähigkeiten und Ansprüchen nicht gerecht werden.

2. Kommunikation & Ansprechbarkeit (vgl. Kapitel 3.1.2)

Die Interaktion mit Robotern als ‚Teammitglieder‘ ändert auch die herkömmliche Kommunikation (im Vergleich zum menschlichen Arbeitspartner). Abläufe und Anweisungen können z. B. nicht im eigentlichen Sinne besprochen werden. In den meisten Anwendungsfällen liegt der Kommunikationsschwerpunkt daher nicht auf verbaler, sondern auf einer non-verbaler Kommunikation mit dem Roboter, die über Gestik, „Mimik“ oder auch die Bewegung selbst vermittelt wird. Es bleibt zu klären, wie eine optimale Kommunikation aussehen könnte und welche Kommunikationsform angemessen ist.

Um den Aspekt der *ethisch-sozialen Implikationen humanoider und zoomorpher Projektionen* genauer zu betrachten, werden für die Analyse folgende Punkte aufgegriffen:

3. Verantwortungsbewusstsein (vgl. Kapitel 3.2.1)

Im Gegensatz zu herkömmlichen Maschinen kann es in der Interaktion mit (insbesondere autonom agierenden) Robotern zu einer Verantwortungsdiffusion kommen, die sich beispielsweise in einer Distanzierung von der Verantwortung beim Menschen äußern kann. In diesem Zusammenhang stellt sich z. B. die Frage, wie menschliche Mitarbeiter die eigene Verantwortung für Anteile der Interaktion in Abhängigkeit von der Gestaltung der MRI empfinden. Auch Zuschreibungen von Verantwortung an den Roboter sind hier relevant.

4. Förderung gewünschter Verhaltensweisen in sozialer Hinsicht (vgl. Kapitel 3.2.1 & 3.2.2)

Es stellen sich Fragen danach, wie sich Interaktionen mit Robotern, die keine sozialen Kompetenzen und Eigenschaften besitzen, entwickeln. So könnte die Zusammenarbeit mit dem ‚Kollegen Roboter‘ zu sozialen Projektionen führen, die die Entscheidungen des Menschen negativ beeinflussen. Mittel- und langfristig sind in diesem Zusammenhang Auswirkungen auf kollegiale Umgangsformen und gesellschaftliche Normen in der Lebens- und Arbeitswelt denkbar. Hier interessiert, welche Eigenschaften der MRI diesen Effekt produzieren bzw. verhindern könnten.

5. Akzeptanz vs. Ablehnung von Robotern (vgl. Kapitel 3.2.2)

Die Akzeptanz bzw. Ablehnung von Robotern ist insofern ethisch relevant, als dass die Beweggründe dieser Einstellungen eher irrationaler Natur und leicht beeinflussbar sind. Darüber hinaus sind auch der Verlust der sozialen Dimension und die Kompensation dieses Verlusts über die Projektion menschlicher Eigenschaften relevant. Eine Ablehnung des Roboters als Interaktionspartner kann z. B. aus einer enttäuschten Erwartung über Fähigkeiten des Roboters entstehen, die wiederum durch seine Gestaltung hervorgerufen werden.

4.2 Studienauswahl

Von ca. 60 gesichteten Studien fließen insgesamt nur 14 Publikationen in die Analyse ein, die den Einfluss verschiedener Gestaltungsmerkmale der MRI auf mindestens einen ethisch-sozialen Aspekt untersucht haben. Grundvoraussetzung um in die Analyse aufgenommen zu werden ist, dass die Studien mindestens ein Gestaltungsmerkmal des Roboters systematisch variieren (mindestens zweifach gestuft), also mindestens zwei verschiedene Roboter bzw. Variationen des gleichen Roboters untersuchen. Studien, die einen Vergleich zwischen Mensch und Roboter anstreben, werden nicht in die Analyse aufgenommen, da dieser Vergleich keine Aussage darüber zulässt, welche Aspekte des Roboters ausschlaggebend für Studienergebnisse sind.

Die Gestaltungsmerkmale des Roboters stellen also die unabhängigen Variablen dar, der Einfluss auf ethisch-soziale Aspekte drückt sich in den abhängigen Variablen aus.

Allerdings stellen tatsächlich nur wenigen Studien einen direkten Bezug zu ethischen Aspekten in der MRI her. D.h. schon die Studienauswahl erfordert eine interpretatorische Leistung, da während der Studienrecherche festgelegt werden musste, welche in den Studien berichteten abhängigen Variablen eine Operationalisierung der in Kapitel 4.1 berichteten ethisch-sozialen Aspekte darstellen (und welche nicht). Tab. 4.1 gibt einen Überblick der einbezogenen Studien, sowie über deren Bezug zu den vorab erläuterten Kategorien und den entsprechenden Operationalisierungen in Form von unabhängigen (UV) und abhängigen Variablen (AV).

Tab. 4.1 In die Analyse eingegangene Studien, unterteilt nach adressierten ethischen Aspekten

Beachte: Die Publikation von Goetz et al. (2003) geht zweimal in die Analyse ein, da mehrere unabhängige Studien berichtet wurden. Die Studie von Briggs & Scheutz (2014) geht auch zweimal ein, allerdings unter Betrachtung unterschiedlicher abhängiger Variablen.

	Studie	Variation UV	AV
ethisch-soziale Qualität der Arbeit	Gombolay, Gutierrez, Clarke, Sturla, & Shah (2015)	Entscheidungsautorität: manuell (Mensch), semi-autonom (Mensch sucht sich seine Aufgabe, Roboter entscheidet Rest), autonom (Roboter entscheidet alles)	objektive Leistungsmaße; Zufriedenheit mit Entscheidungsautonomie
	Instrumentalisierung vs. Selbstbestimmung Goetz, Kiesler & Powers (2003) Studie II & III	Studie II: verbale Kommunikation des Roboters: verspielt vs. ernsthaft Studie III: verbale Kommunikation des Roboters: verspielt vs. ernsthaft Aufgabe: anstrengend (Sport) vs. lustig/kreativ (Süßigkeiten-Rezepte ausdenken)	Studie II: Dauer, die Probanden eine sportliche Aktivität ausüben nach Aufforderung des Roboters Studie III: Dauer, die Probanden eine sportliche Aktivität ausüben nach Aufforderung des Roboters bzw. Dauer, die Probanden an einer spaßmachenden Aufgabe arbeiten nach Aufforderung des Roboters

	Studie	Variation UV	AV
ethisch-soziale Qualität der Arbeit	Breazeal, Kidd, Thomaz, Hoffman & Berlin (2005)	Kommunikationsform Roboter zu Mensch: EXP explizit non-verbal (z.B. Kopfnicken als Bestätigung) vs. IMP+EXP implizit+explizit non-verbal (Kopfnicken plus Aufmerksamkeitsausrichtung auf bestimmte Objekte)	Verständlichkeit/ Transparenz der Roboter-"intention"; wahrgenommene Verantwortung für Aufgabe; Effizienz in Aufgabenbearbeitung; Fehlerrobustheit in Aufgabenbearbeitung
	Stubbs, Hinds, & Wettergreen (2007)	Autonomie des Roboters: niedrig, mittel, hoch	Etablierung eines "Common Ground" zw. Roboter und Mensch
	Kuz, Mayer, Müller & Schlick (2013)	Bewegungsverhalten des Roboters: konventionell (funktional) vs. anthropomorph	Voraussage der Zielposition der Bewegung des Roboters (ja/nein); Zielgenauigkeit der Voraussage
	Mayer, Kuz & Schlick (2013)	Profil der Bewegungsgeschwindigkeit des Roboters: konventionell vs. anthropomorph	Voraussage der Zielposition der Bewegung des Roboters (ja/nein); Zielgenauigkeit der Voraussage
	Riek, Rabinowitch, Bremner, Pipe, Fraser & Robinson (2010)	(Unter anderem) Profil der Bewegungsgeschwindigkeit des Roboters: geschmeidig (menschähnlich) vs. abrupt (maschinenähnlich)	Reaktionszeit auf Bewegung des Roboters; Bewegungsverständnis (Information der Bewegung)

	Studie	Variation UV	AV	
Projektion lebendiger Eigenschaften auf Roboter		Autonomie des Roboters: konfrontatives Verhalten (Roboter protestiert gegen "unmoralische" Entscheidung den Turm zu zerstören") vs. non-konfrontativ (Roboter nimmt Direktiven einfach an)	Zerstörung des Turms (Ja/nein)	
	Verantwortungsbewusstsein	Briggs, Gessell, Dunlap & Scheutz (2014)	Robotermorphologie: humanoider Roboter vs. funktionaler Roboter	Akzeptanz von Autonomie & moralischer Vorstellung von Roboter: Befehle geben obwohl Roboter protestiert
		Kim & Hinds (2006)	Autonomie des Roboters: hoch vs. niedrig Transparenz des Roboters: hoch vs. niedrig	Einschätzung des eigenen Verdienstes und dem des Roboters an erfolgreicher Aufgabenerfüllung; Attribuierte Schuld: selbst und Roboter
		Hinds, Roberts & Jones (2004)	Menschähnlichkeit: Mensch, humanoider Roboter, funktionaler Roboter Roboterstatus: Untergebener, Kollege, Überwacher	wahrgenommene Verantwortung für Aufgabe

	Studie	Variation UV	AV
Projektion lebendiger Eigenschaften auf Roboter	Parise, Kiesler, Sproull & Waters (1996)	Morphologie eines Avatars (stellvertretend für Roboter): Person, Hund, Cartoon-Hund, echter Mensch (Video)	Kooperationsverhalten
	Förderung gewünschter Verhaltensweisen Riek, Rabinowitch, Chakrabarti & Robinson (2009)	Robotermorphologie: 5 Stufen von rein funktional (Roomba) über anthropomorph bis hin zu echtem Jungen (als Kontrollvariable)	Empathie

	Studie	Variation UV	AV
Projektion lebendiger Eigenschaften auf Roboter	Darling, Nandy & Breazeal (2015)	<i>Framing:</i> Roboter mit und ohne Hintergrundgeschichte Roboterbewegung: ja/nein	Weigerung/Zögern, den Roboter zu zerstören (Hexbug)
	Akzeptanz vs. Ablehnung von Robotern Briggs, G., & Scheutz, M. (2014)	Autonomie des Roboters: konfrontatives Verhalten (Roboter protestiert gegen "unmoralische" Entscheidung den Turm zu zerstören") vs. non-konfrontativ (Roboter nimmt Direktiven einfach an)	Wohlbefinden der VPn; Einschätzung des Begriffsvermögens des Roboters
	Goetz, Kiesler & Powers (2003) Studie I	Menschähnlichkeit des Roboters: humanoid, weniger humanoid, technisch/maschinell	Häufigkeit, mit der ein bestimmter Roboter für Aufgaben eines Serviceroboters gewählt wurde

4.3 Systematisierung der Einflussgrößen anhand der MRI-Taxonomie

4.3.1 Die MRI-Taxonomie

Im Gegensatz zu anderen Klassifikationsansätzen (Scholtz, 2002; Yanco & Drury, 2002, 2004) fokussiert die von Onnasch et al. (2016) entwickelte Taxonomie explizit die Interaktionskomponente zwischen Roboter und Mensch. Dabei genügt sie dem Anspruch einer möglichst hohen Generalisierbarkeit (Anwendung in verschiedensten Domänen auf unterschiedliche Robotertypen) und ermöglicht eine Interaktionscharakterisierung von Mensch und Roboter mit hohem Detailgrad. Die Klassifikation unterteilt sich in drei verschiedene Cluster: Interaktion, Roboter und Team. Diesen drei Kategorien sind weitere Beschreibungsmerkmale zugeordnet, die in Abb. 4.1 dargestellt sind.

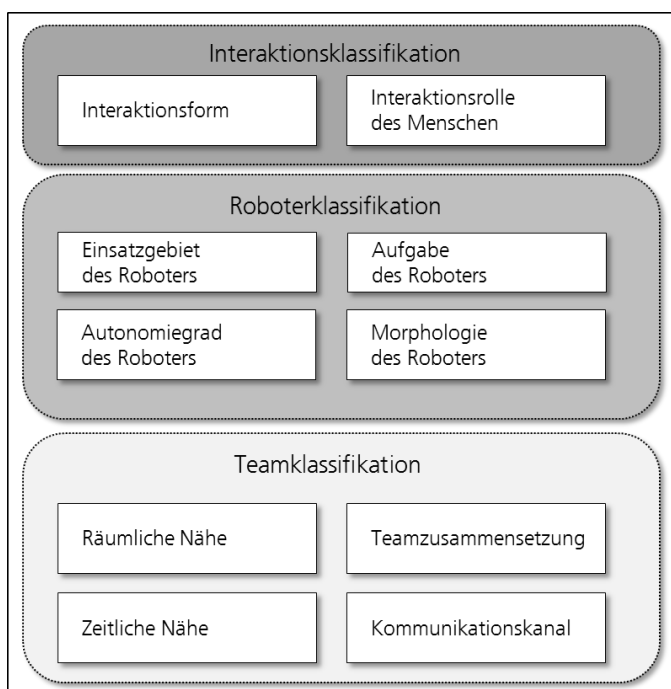


Abb. 4.1 Schematische Darstellung der MRI-Taxonomie: Klassifikationscluster und Beschreibungsmerkmale

Um die Beschreibung konkreter MR-Szenarien zu vereinfachen, wurde die Taxonomie in ein Canvas-Format übertragen. Die Taxonomie erlaubt einen strukturierten Vergleich verschiedener MR-Systeme, so dass z. B. die Identifikation übergeordneter Einflussgrößen auf die MRI möglich wird. Auf diese Weise kann der Einfluss bestimmter Gestaltungsmerkmale des Roboters auf typische Human-Factors-Aspekte wie z. B. Vertrauen, Akzeptanz, Situationsbewusstsein oder mentale Beanspruchung einer systematischen (Meta-)Analyse unterzogen werden.

Im vorliegenden Projekt wird der Bezug der Taxonomie auf ethische Aspekte erweitert. Ziel ist, konkrete Dimensionen zu identifizieren, die besonders relevant für eine gelungene MR-Interaktion aus ethisch-soziologischem Blickwinkel sind.

Abb. 4.2 zeigt die Canvas-Darstellung der bestehenden Taxonomie. Die drei Klassifikationscluster sind durch farbliche Abstufungen gekennzeichnet (Interaktionsklassifi-

kation = dunkelgrau, Roboterklassifikation = mittelgrau, Teamklassifikation = hellgrau).

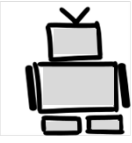
<u>Illustration und konkrete Beschreibung des Roboters</u> 		<u>Interaktionsform Mensch ↔ Roboter</u> Kollaboration Kooperation Ko-Existenz	<u>Interaktionsrolle des Menschen</u> Supervisor Operateur Kollaborateur Kooperateur Nicht-Beteiligter
		<u>Aufgabe des Roboters</u> Informationsaustausch Präzision Entlastung Transport Manipulation	
<u>Kommunikationskanal</u> M → R Elektronisch Mechanisch Akustisch Optisch R → M Mechanisch Akustisch Visuell		<u>Morphologie des Roboters</u> Humanoid Zoomorph Funktional	<u>Autonomiegrad des Roboters</u> Infoaufnahme ———— Infoverarbeitung ———— Entscheidungsfindung ———— Handlungsausführung ———— hoch
		<u>Räumliche Nähe</u> Berührend Annähernd Führend Vorbeigehend Vermeidend Ferngesteuert	
		<u>Teamzusammensetzung Verhältnis Mensch - Roboter</u> $N_M = N_R$ $N_M > N_R$ $N_M < N_R$	

Abb. 4.2 Canvas-Darstellung der MRI-Taxonomie

Die Darstellung dient als Vorlage zur Charakterisierung konkreter Interaktionsszenarien, weshalb pro Variable alle möglichen Ausprägungen aufgeführt sind. Bei der Anwendung der Taxonomie auf konkrete Beispiele sollte pro Variable (mindestens) eine Ausprägung gewählt werden.

4.3.2 Anwendung der Taxonomie auf unabhängige Variablen der Studien

Die unabhängigen Variablen (UV) der in die Analyse eingehenden Studien beziehen sich auf verschiedene Aspekte der drei Klassifikationscluster der Taxonomie. Für drei der insgesamt 14 Studien werden die unabhängigen Variablen als Variationen der **Interaktionsrolle des Menschen** kategorisiert (Goetz et al., 2003; Gombolay et al., 2015; Hinds et al., 2004).

In der Studie von Goetz et al. (2003) wird als UV die Kommunikationsform (Sprache) des Roboters einmal als spielerisch und einmal als ernsthaft operationalisiert. In der spielerischen Variante kommuniziert der (humanoide) Roboter eher kumpelhaft und lustig. In der ernsthaften Variante wird ein Befehlston angewandt, der Roboter sagt dem Probanden, was dieser zu tun hat. Dadurch ändert sich das Verhältnis der Interaktionspartner von einem eher ebenbürtigen Niveau hin zu einem Szenario, in dem der Roboter eine führende Funktion einnimmt. Im Sinne der MRI-Taxonomie ändert sich die Rolle des Probanden von einem Kollaborateur (gemeinsame Aufgabenbearbeitung und Zielerreichung) hin zu der eines ‚untergebenen‘ Mitarbeiters.

Hier wird eine ‚Führungsrolle‘ des Roboters suggeriert, die aus den in Kapitel 3.1.1 genannten Gründen problematisch ist.

Auch in der Studie von Gombolay et al. (2015) wird dieser Rollenwechsel vollzogen. In diesem Fall wird als UV die Entscheidungsautorität dreifach abgestuft: Von einer rein „manuellen“ Entscheidung, wie nachfolgende Aufgaben zwischen Mensch und Roboter verteilt werden sollten, über eine gemischte Entscheidung hin zu einer reinen Roboterentscheidung. Während der Mensch im ersten Fall also die Funktion eines Supervisors einnimmt, im zweiten Fall die eines Kollaborateurs, ändert sich die Situation im letzten Fall, da hier wieder der Roboter die führende Funktion in der Aufgabenplanung einnimmt.

Bei Hinds et al. (2004) ist die Rollenklassifikation direkter auf die Taxonomiedimension „Interaktionsrolle“ des Menschen zu beziehen, da hier systematisch zwischen dem Roboter als Untergebenem (Mensch als Supervisor), als gleichwertigem Mitarbeiter (Mensch als Kollaborateur) und führendem Mitglied unterschieden wird (operationalisiert über framing).

In drei weiteren Studien werden die UV als Operationalisierung des **Autonomiegrads des Roboters** gewertet (Briggs & Scheutz, 2014; Kim & Hinds, 2006; Stubbs et al., 2007). Bei Briggs & Scheutz (2014) wird die Autonomie über die Gegenwehr des Roboters operationalisiert. Während ein Roboter in einer Bedingung Befehle einfach annimmt und ausführt, zeigt der Roboter in einer anderen Bedingung ein Protestverhalten, da er den Befehl/ die Entscheidung des Probanden ungerecht findet (es geht darum, einen Turm zu zerstören, den der Roboter bzw. ein anderer Roboter zuvor gebaut hat). Insistiert der Proband, wird die Aktion vom Roboter ausgeführt. Diese Operationalisierung entspricht einer Variation des Autonomiegrads auf der Entscheidungsfindungsebene (gering vs. hoch).

Auch Kim & Hinds (2006) variieren zwischen einer hohen Autonomie und einer geringen Autonomie, indem sie die Entscheidungsfindung entweder dem Roboter überlassen oder aber dem Menschen. Dies entspricht wieder einer Variation von gering zu hoch auf der Entscheidungsfindungsebene des Autonomiegrads.

Stubbs et al. (2007) operationalisieren die drei Faktorstufen ihrer UV über verschiedene Ebenen des Autonomiegrads. Während in der niedrigen Bedingung der Roboter nur wenige eigenständige Aktionen ausführen kann und die meiste Zeit von einem Operateur ferngesteuert werden muss, ermöglicht die moderate Autonomie dem Roboter bereits eigenständige Entscheidungen und Aktionen, die allerdings noch immer von extern gegebenen Informationen abhängig sind. In der höchsten Autonomiestufe agiert der Roboter weitestgehend autonom. Dies entspricht Variationen auf den Stufen der Informationsverarbeitung, Entscheidungsfindung und Handlungsausführung.

Die meisten Studien variieren allerdings die **Morphologie des Roboters**. Insgesamt zehn Studien beschäftigen sich mit dem Einfluss einer anthropomorphen (oder auch zoomorphen, s. Darling et al., 2015; Parise et al., 1996) Gestaltung von Robotern. Sechs der Studien variieren die Morphologie über das Aussehen des Roboters unter Einbeziehung der Stufen funktional, zoomorph und humanoid, wobei nicht jede Studie alle drei Faktorstufen betrachtet (Briggs et al., 2014; Goetz et al., 2003; Hinds et al., 2004; Kahn et al., 2012; Parise et al., 1996; Riek et al., 2009).

Andere Studien operationalisieren verschiedene Morphologieausprägungen z. B. über das Bewegungsverhalten des Roboters (Trajektorien und Geschwindigkeit: (Kuz et al., 2013; Mayer et al., 2013; Riek et al., 2010) über das Kommunikationsverhalten (Breazeal, Kidd, Thomaz, Hoffman & Berlin, 2005) oder aber über den Kontext, der den Probanden über Vorgeschichten bzw. Beschreibungen des Roboters gegeben wird (*framing*, Darling et al., 2015).

4.3.3 Erweiterung der MRI-Taxonomie

Drei Merkmalsbeschreibungen der MRI-Taxonomie (Onnasch et al., 2016) wurden auf die ausgewählten Studien angewendet: die Interaktionsrolle des Menschen, der Autonomiegrad des Roboters und die Morphologie. Während aufgrund der Anwendung der Taxonomie auf diese Studien kein Änderungsbedarf für die Kategorie Autonomiegrad ersichtlich wurde, weisen die Studien darauf hin, dass auf den ersten Blick für die Interaktionsrolle des Menschen sowie für die Morphologie eine Überarbeitung bzw. Erweiterung der Merkmalsausprägungen notwendig erscheint.

Alle hier berichteten Studien, die sich mit der Interaktionsrolle des Menschen auseinandersetzen, haben diese auch dahingehend variiert, dass der Mensch dem Roboter ‚untergeordnet‘ ist, dem Roboter also eine Art ‚Führungsrolle‘ zugeschrieben wird. Diese Rollenverschiebung ist in der bisherigen Merkmalsbeschreibung nicht vorgesehen. Auch wenn den Probanden in den Studien eine ‚Führungsrolle‘ von Robotern suggeriert wurde, muss daran festgehalten werden, dass Roboter faktisch keine Führungsrolle im starken Sinne einnehmen können, da sie keine personale Autonomie besitzen und keine Verantwortung übernehmen können (vgl. Kap. 3.1.1). Die Suggestion einer ‚Autorität‘ von Robotern über Menschen ist folglich unangemessen und problematisch. Eine führende Funktion eines Roboters kann nur entweder eine Anleitung im Sinne eines Taktgebers oder eine Handlungsvorgabe als Ergebnis informationsverarbeitender Prozesse darstellen. Entsprechende Interaktionen treten als Varianten von Kooperationen oder Kollaborationen auf. Deshalb wird auf eine Erweiterung dieser Kategorie innerhalb der Taxonomie verzichtet.

Die Morphologie ist zentrales Thema der meisten hier betrachteten Studien. Diese wurde nicht nur über die äußere Gestaltung des Roboters variiert, sondern auch über die Kommunikation, über *framing* oder auch über die Trajektorienwahl und -geschwindigkeit. Innerhalb der MRI-Taxonomie wird die Merkmalsbeschreibung Morphologie in drei Ausprägungen unterteilt: anthropomorph, zoomorph und funktional. Diese relativ generische Beschreibung wird den vielfältigen Möglichkeiten, Morphologie zu operationalisieren, nicht gerecht. Daher wird dieses Merkmal der MRI in der Taxonomie entsprechend aktualisiert: Während die Ausprägungen weiterhin dreifach gestuft bleiben, wird die Taxonomie erweitert, indem zwischen **Kontextmorphologie** (z. B. über *framing*), **Gestaltmorphologie** (sichtbares äußeres Design des Roboters), **Kommunikationsmorphologie** (z. B. Sprache, non-verbale Kommunikation, Gestik) und **Bewegungsmorphologie** unterschieden wird.

Abb. 4.3 zeigt die MRI-Taxonomie mit den entsprechenden Erweiterungen.


<u>Illustration und konkrete Beschreibung des Roboters</u>		<u>Interaktionsform</u> Mensch ↔ Roboter	<u>Interaktionsrolle des Menschen</u>
	<u>Aufgabe des Roboters</u> Informationsaustausch Präzision Entlastung Transport Manipulation	<u>Einsatzgebiet des Roboters</u> Industrie Kommerz. Service Pers. Service	Supervisor Operateur Kollaborateur Kooperationspartner Nicht-Beteiligter
	<u>Räumliche Nähe</u> Berührend Annähernd Führend Vorbeigehend Vermeidend Ferngesteuert	<u>Zeitliche Nähe</u> Synchron Asynchron	<u>Teamzusammensetzung</u> <u>Verhältnis Mensch - Roboter</u> $N_M = N_R$ $N_M > N_R$ $N_M < N_R$

Abb. 4.3 Canvas-Darstellung der MRI-Taxonomie mit Erweiterung der Dimension "Morphologie des Roboters"

4.4 Einfluss der Interaktionsrolle auf ethisch-soziale Aspekte

Die Analyse der Studien ergibt, dass die Ausgestaltung der Interaktionsrolle insbesondere in Bezug zu einer potentiellen Instrumentalisierung des Menschen im Arbeitskontext eine entscheidende Rolle spielt. Grundsätzlich akzeptieren Personen Roboter sowohl in gleichgestellten Positionen als auch in führenden Funktionen. So zeigen Gombolay et al. (2015), dass Personen bei einer Planungsaufgabe sogar eine führende Funktion des Roboters und dessen Entscheidungsfindung bevorzugen. Dies ist mit der Art der Aufgabe zu begründen. Die Rollenverteilung muss der Aufgabe angemessen sein und sowohl die Fähigkeiten des Menschen als auch die des Roboters widerspiegeln (vgl. Kap. 3.1.1). Wenn Aufgaben z. B. Optimierung und Effizienzsteigerung zum Ziel haben, werden Roboter als fähiger wahrgenommen, selbst Anleitungen durch den Roboter werden hier vom Menschen akzeptiert.

Je stärker der Mensch in einer gemeinsamen Aufgabe die Interaktion führt, desto höher ist auch das Verantwortungsgefühl für die Gesamtaufgabe von Roboter und Mensch. Da Roboter de facto keine Verantwortung übernehmen können (vgl. Kapitel 2.2), sollte daher die Suggestion einer führenden Funktion des Roboters nicht zu stark ausgeprägt sein, so dass der Mensch dieses Bewusstsein aufrechterhält. Zusammenfassend ist festzuhalten, dass die Interaktionsrolle immer entsprechend des Aufgabenprofils gewählt werden sollte, eine grundsätzliche Empfehlung zur Rollenverteilung ohne Aufgabenkontext erscheint nicht zielführend.

Die Ergebnisse der Analyse der einzelnen Studien sind in Tab. 4.2 nochmals zusammengefasst dargestellt.

Tab. 4.2 Einfluss der Interaktionsrolle des Menschen auf ethisch-soziale Aspekte der MRI

Beschreibungsmerkmal Interaktionsrolle des Menschen			
Studie	UV-Variation	Entspricht in Taxonomie	Interpretation der Ergebnisse
Goetz, Kiesler, & Powers, 2003	Kommunikationsform spielerisch vs. ernsthaft	Interaktionsrolle (Kollaborateur vs. Untergebener)	Probanden führen Aufgaben länger aus wenn sie untergeordnet sind Interaktionsrolle muss Aufgabenprofil entsprechen (soziale Aufgaben eher kollaborativ ausgestaltet, bei rein leistungsorientierten Aufgaben Roboter in Führungsrolle)
Gombolay, Gutierrez, Clarke, Sturla, & Shah, 2015;	Entscheidungsautorität: Mensch, gemischt, Roboter	Interaktionsrolle (Supervisor, Kollaborateur, Untergebener)	Probanden bevorzugen semi-autonome und autonome Entscheidungsfindung vor manueller Entscheidung, sind zufriedener wenn Roboter Aufgaben allokiert → Aufgabenprofil quantitativ leistungsorientiert, entspricht stärker Fähigkeiten des Roboters, deshalb hohe Akzeptanz
Hinds, Roberts, & Jones, 2004	Supervisor, Kollaborateur, Untergebener	Interaktionsrolle (Supervisor, Kollaborateur, Untergebener)	Je stärker die Führungsrolle des Menschen, desto höher das Verantwortungsgefühl für Gesamtaufgabe (Roboter + Mensch)

4.5 Einfluss der Autonomie auf ethisch-soziale Aspekte

Der Autonomiegrad des Roboters beeinflusst wesentlich die Akzeptanz des Roboters als auch die wahrgenommene Verantwortung des Menschen für die Gesamtaufgabe. So zeigen die Studien von Briggs & Scheutz (2014) und Kim & Hinds (2006), dass Probanden mit steigender Autonomie des Roboters, insbesondere auf der Ebene der Entscheidungsfindung, zunehmend passiver werden. D. h. die Einflussnahme auf die Aufgabenausführung durch den Roboter steigt mit zunehmendem Autonomiegrad bei gleichzeitiger Abnahme des subjektiven Verantwortungsgefühls der Probanden. Auch die Akzeptanz des Roboters als „Person“ ist wesentlich durch dessen Autonomie beeinflusst. So konnten Briggs & Scheutz (2014) zeigen, dass sich Probanden unwohler fühlen, wenn sie einem autonomen Roboter widersprechen müssen – im Gegensatz zum Widerspruch gegen einen weniger autonomen Roboter. Dieses Unbehagen ist nicht mit objektiv rationalen Gründen zu erklären, da es sich auch bei einem autonomen Roboter immer noch um ein technisches Artefakt handelt. Diese affektive Reaktion und die zunehmende Beeinflussbarkeit des Menschen in seinen Entscheidungen durch autonome Roboter ist aus ethischer Perspektive kritisch zu hinterfragen, da diese Mechanismen auch potentielle Manipulationsmöglichkeiten im Sinne einer Instrumentalisierung des Menschen durch den Einsatz von Robotern eröffnen.

Deshalb wird mit steigendem Autonomiegrad des Roboters auch die *Transparenz* ein zunehmend wichtiges Thema. Transparenz führt nicht nur zu einer besseren Nach-

vollziehbarkeit von Entscheidungen und Aktionen des Roboters (Stubbs et al., 2007), sondern Transparenz im Sinne von Aufklärung über den Roboter und dessen Funktionsweise kann auch als Prävention gegen eine zunehmend empfundene Instrumentalisierung und Abhängigkeit des Menschen wirken.

Die Ergebnisse der Analyse der einzelnen Studien sind in Tab. 4.3 nochmals zusammengefasst dargestellt.

Tab. 4.3 Einfluss der Roboterautonomie auf ethisch-soziale Aspekte der MRI

Beschreibungsmerkmal Autonomiegrad des Roboters			
Studie	UV-Variation	Entspricht in Taxonomie	Interpretation der Ergebnisse
Briggs & Scheutz, 2014	Einflussnahme auf Entscheidung des Menschen: keine vs. Einspruch	Autonomiegrad Entscheidungsfindung gering vs. hoch	Je höher die Autonomie des Roboters desto wahrscheinlicher lassen sich Menschen in ihrer Entscheidung beeinflussen. Je höher die Autonomie des Roboters desto unwohler fühlen sich Menschen wenn sie gegen den Roboter entscheiden.
Kim & Hinds, 2006	Autonomie niedrig vs. hoch	Autonomiegrad Entscheidungsfindung gering vs. hoch	Je höher der Autonomiegrad, desto weniger verantwortlich fühlen sich Personen für Aufgabenerfüllung. Je höher der Autonomiegrad, desto geringer schätzen Personen ihren eigenen Anteil zur erfolgreichen Aufgabenerfüllung ein.
Stubbs, Hinds, & Wettergreen, 2007	Autonomie: niedrig, mittel, hoch	Autonomiegrad Informationsverarbeitung, Entscheidungsfindung, Handlungsausführung niedrig, mittel, hoch	Je höher die Autonomie, desto wichtiger wird Kenntnis über Gründe der autonomen Entscheidungen und Aktionen (Transparenz).

4.6 Einfluss der Morphologie auf ethisch-soziale Aspekte

4.6.1 Gestaltmorphologie

Die Studienanalyse legt nahe, dass über die äußere Gestaltung des Roboters gewünschte Verhaltensweisen auf Seiten des Menschen gefördert werden können. So zeigen die Ergebnisse der Studien von Parise et al. (1996) und Riek et al. (2009), dass mit einem steigenden Ähnlichkeitsgrad des Roboters zum Menschen sowohl die Empathie als auch die Kooperationsbereitschaft steigen. Diese Bereitschaft geht sogar so weit, dass Menschen auf ihren eigenen Maximalnutzen zu Gunsten des Roboters verzichten. Die Ergebnisse legen insofern auch nahe, dass diese positive Einflussnahme über die Robotergestalt auch ein großes Manipulationspotential des interagierenden Menschen aufweist. Deshalb sollte eine anthropomorphe Robotergestaltung im Arbeitskontext immer so gering wie möglich ausfallen und nach den Erfordernissen der gemeinsam auszuführenden Aufgabe gewählt werden (s. Goetz et al., 2003). Dies ist auch in Bezug zur wahrgenommenen Aufgabenverantwortung des Menschen in der MRI ratsam, da auch mit steigendem Anthropomorphismus die wahrgenommene Verantwortung des Menschen für die Gesamtaufgabe abnimmt (Hinds et al., 2004).

Die Ergebnisse der Analyse der einzelnen Studien sind in Tab. 4.4 nochmals zusammengefasst dargestellt.

Tab. 4.4 Einfluss der Gestaltmorphologie auf ethisch-soziale Aspekte der MRI

Beschreibungsmerkmal Morphologie des Roboters - Gestaltmorphologie			
Studie	UV-Variation	Entspricht in Taxonomie	Interpretation der Ergebnisse
Briggs, Gessell, Dunlap, & Scheutz, 2014	funktional vs. humanoid	Gestalt funktional vs. humanoid	Kein Einfluss, jedoch ist festzuhalten, dass auch ein übertriebenes emotionales Mimikri des Roboters zu Unglaubwürdigkeit führt. Ein rein quantitatives "Mehr" an menschlichen Ausdrücken hat nicht zwangsläufig Einfluss auf die wahrgenommene Autonomie des Roboters.
Goetz et al., 2003	humanoid, weniger humanoid, funktional	Gestalt humanoid, weniger humanoid, funktional	Morphologie muss Aufgabenprofil entsprechen (soziale Aufgaben eher humanoid, ernsthafte/ objektivierbare Aufgaben eher funktional)
Hinds et al., 2004	Mensch (Kontrollbedingung), menschlicher Roboter, sehr technisch wirkender Roboter	Gestalt humanoid vs. funktional	Je anthropomorpher der Roboter gestaltet ist, desto weniger Verantwortung empfindet der Mensch für Gesamtaufgabe (Mensch + Roboter)
Parise et al., 1996	Humanoider Agent, zoomorpher Agent, Comicähnlicher zoomorpher Agent, Mensch (Kontrollbedingung)	Gestalt humanoid vs. zoomorph	Kooperationsverhalten des Menschen steigt mit zunehmendem Ähnlichkeitsgrad eines technischen Agenten (funktional – zoomorph – humanoid). Kooperationsbereitschaft bedeutet hier sogar auf den eigenen Maximalnutzen zu verzichten.
Riek, Rabinowitch, Chakrabarti, & Robinson, 2009	5 Stufen von rein funktional (Roomba) über anthropomorph bis hin zu echtem Jungen (als Kontrollvariable)	Gestalt funktional vs. humanoid	Mit steigendem Ähnlichkeitsgrad des Roboters steigt auch die Empathie des Menschen

4.6.2 Bewegungsmorphologie

Die Bewegungsmorphologie des Roboters ist ein wichtiger Aspekt der non-verbalen Kommunikation des Roboters mit dem Menschen. Durch eine humanoide Wahl der Bewegungstrajektorien und -geschwindigkeit steigt die Koordinationsleistung des Menschen in der Interaktion mit dem Roboter. So zeigen die Studien von Kuz et al. (2013) und Mayer et al. (2013), dass durch ein anthropomorphes Bewegungsprofil (Trajektorien und Geschwindigkeit) die Intentionserkennung der Roboterhandlungen für den Menschen deutlich vereinfacht wird (im Gegensatz zu einer rein funktionalen / effizienten Bewegungsgestaltung). Dies ist eine grundlegende Voraussetzung für eine optimale Interaktionskoordination. Studien belegen, dass hierfür das motorische System eine zentrale Rolle spielt und weniger die tatsächliche verbale Kommunikation (Rizzolatti & Craighero, 2004). Wenn Personen Handlungen beobachten, werden diese mental repräsentiert, so dass eigene Handlungen besser angepasst werden können (Atmaca, Sebanz, Prinz, & Knoblich, 2008; Sebanz, Knoblich, & Prinz, 2003, 2005). Allerdings werden Handlungen des Gegenübers nur dann mental repräsentiert, wenn das Gegenüber als intentional handelnd wahrgenommen wird (Sebanz & Knoblich, 2009). Diese Voraussetzung scheint durch eine anthropomorphe Bewegungsgestaltung erfüllt zu werden.

Allerdings ist anzumerken, dass die Befunde von Riek et al. (2010) in die gegenteilige Richtung weisen. Die Befunde von Kuz et al. (2013) und Mayer et al. (2013) werden trotzdem höher gewichtet und entsprechend interpretiert, da Riek et al. (2010) einzig die Schnelligkeit, mit der Roboterhandlungen identifiziert werden, evaluiert haben, nicht aber, ob Handlungen wirklich korrekt identifiziert wurden. Außerdem han-

delte es sich um soziale Interaktionsgesten, bei Kuz und Kollegen wurde ein industrieller Kontext gewählt.

Die Ergebnisse der Analyse der einzelnen Studien sind in Tab. 4.5 nochmals zusammengefasst dargestellt.

Tab. 4.5 Einfluss der Bewegungsmorphologie auf ethisch-soziale Aspekte der MRI

Beschreibungsmerkmal Morphologie des Roboters - Bewegungsmorphologie			
Studie	UV-Variation	Entspricht in Taxonomie	Interpretation der Ergebnisse
Kuz, Mayer, Müller, & Schlick, 2013	Bewegungsprofil: konventionell (funktional) vs. antropomorph	Bewegung funktional vs. humanoid	Je anthropomorpher die Bewegung, desto höher die Wahrscheinlichkeit einer Handlungsvorhersage. Je anthropomorpher die Bewegung, desto deutlicher ist die Handlungsintention/ desto einfacher die Intentionserkennung.
Mayer, Kuz, & Schlick, 2013	Bewegungsschnelligkeit: konventionell (funktional) vs. antropomorph	Bewegung funktional vs. humanoid	Je anthropomorpher die Bewegung, desto höher die Wahrscheinlichkeit einer Handlungsvorhersage. Je anthropomorpher die Bewegung, desto deutlicher ist Handlungsintention/ desto einfacher die Intentionserkennung.
Riek u. a., 2010	Bewegungsschnelligkeit: geschmeidig (menschähnlich) vs. abrupt (maschinenähnlich)	Bewegung funktional vs. humanoid	Je maschinenartiger die Bewegung, desto schneller sind Handlungen identifizierbar.

4.6.3 Kommunikationsmorphologie

Zur Kommunikationsmorphologie liegt nur eine Studie vor, die sich mit dem Einfluss non-verbaler Kommunikation des Roboters auf Effizienz und Fehlerrobustheit der Leistung in der MRI beschäftigte (Breazeal et al., 2005). Die non-verbale Kommunikation erfolgte entweder über explizite non-verbale Hinweise (Bedingung EXP), z. B. Kopfnicken eines zoomorphen Roboters als Bestätigung, oder eine Kombination aus impliziten und expliziten Hinweisen (Bedingung EXP+IMP), z. B. Kopfnicken in Kombination mit einer Aufmerksamkeitsausrichtung auf ein bestimmtes Objekt. Die Probanden sollten den Roboter (s. Abb. 4.4), der mit Augen, Hände und Ohren ausgestattet war, über Sprache und Gesten anlernen, drei verschiedenfarbige Knöpfe zu drücken. Anschließend sollten sie dem Roboter befehlen, bestimmte Knöpfe zu drücken. Als abhängige Variablen wurden die Verständlichkeit / Transparenz der Handlungsabsicht, die Effizienz der Aufgabenbearbeitung und die Fehlerrobustheit der Aufgabenbearbeitung betrachtet. Die Ergebnisse zeigen, dass Probanden der EXP+IMP Bedingung die Handlungen des Roboters bzw. den Roboter selbst besser verstehen und auch ein besseres mentales Modell des Roboters aufbauen können als Probanden der EXP-Bedingung. Darüber hinaus brauchen sie weniger Zeit zur Aufgabenerledigung, Roboter und Mensch machen weniger Fehler (falschen Knopf drücken) und die Fehler des Roboters werden schneller detektiert und können somit schneller behoben bzw. verhindert werden.



Abb. 4.4 Versuchsaufbau der Studie von Breazeal und Kollegen (2005)

Allerdings lässt die Studie auch einige Kritikpunkte zu. Dass Fehler in der EXP+IMP-Bedingung schneller detektiert wurden ist nicht verwunderlich, da der Roboter in diesem Fall immer vor der Handlung auf den Knopf schaute, um den es ging. Wenn er also vor der Handlung auf einen falschen Knopf schaute, dann war dies sofort ersichtlich. In der EXP-Bedingung schaute der Roboter bis zur Handlungsausführung nur stur geradeaus. Es handelt sich hier also um einen Vorteil im Sinne eines höheren Informationsgehalts (zusätzlicher Hinweis vor der eigentlichen Handlung), dieser wurde in den Bedingungen nicht konstant gehalten.

Darüber hinaus wurde nicht systematisch zwischen einer nur EXP-Bedingung und einer nur IMP-Bedingung unterschieden. Dies wäre aber besonders interessant, da es in der Robotergestaltung auch immer darum gehen sollte, welche Anforderungen minimal einzuhalten sind, um maximale Vorteile zu generieren. So wäre vorstellbar, dass gleiche oder zumindest ähnliche Ergebnisse in einer IMP-Bedingung hätten erzielt werden können. Als letzten Kritikpunkt ist anzumerken, dass die Verhaltensdaten nur deskriptiv ausgewertet werden konnten, da die Roboterfehler nicht systematisch „kontrolliert“ wurden, sondern zufällig auftraten. D. h., ob es sich bei den Ergebnissen um statistisch und praktisch relevante Befunde handelt, kann nicht mit Sicherheit gesagt werden.

Angesichts der Studienbefunde sowie vor dem Hintergrund der genannten Kritikpunkte wird folgende Hypothese zur Kommunikationsmorphologie aufgestellt, die in Zukunft experimentell überprüft werden müsste:

Implizite non-verbale Kommunikation hat den höchsten dynamischen Informationsgehalt, erfordert in der Interpretation am wenigsten kognitive Ressourcen und ist kulturübergreifend eindeutig. Je nach Bedarf kann sie mit anderen Kommunikationsformen erweitert werden, z. B. bei geforderten Standards einer expliziten Bestätigung.

Auch wenn aus dieser Hypothese und der Studie keine unmittelbaren Schlussfolgerungen über ethische Aspekte der Kommunikation zwischen Mensch und Roboter in der MRI gezogen werden können, so ist doch anzunehmen, dass mit einer impliziten non-verbale Kommunikation (insbesondere im Vergleich mit einer simulierten verbalen Kommunikation) eine flüssige Interaktion unterstützt werden kann, die nicht übermäßig humanoide Assoziationen und Projektionen mobilisiert. Die Zuschreibung von intentionalen Zuständen beispielsweise durch eine Fokussierung von ‚Roboter-Augen‘ auf ein Objekt orientiert sich zwar offensichtlich an intuitiven Praktiken der

Kommunikation zwischen Menschen oder auch zwischen Menschen und Tieren; es ist aber bisher nicht ersichtlich, ob und wie in diesem Zusammenhang Probleme wie Instrumentalisierung oder Verantwortungsdiffusion auftreten können. Insbesondere im Zusammenhang mit humanoiden und zoomorphen Projektionen wären hier allerdings weitere Untersuchungen erforderlich, auf deren Grundlage eine genaue Einschätzung der ethisch-sozialen Relevanz möglich wäre.

4.6.4 Kontextmorphologie

Zum Einfluss des Kontextes wurde in der Studienrecherche nur eine Studie identifiziert, jedoch mit starken Befunden. Darling et al. (2015) untersuchten den Einfluss von framing auf die Bereitschaft von Probanden, einen kleinen Roboter mit einem Hammer zu zerstören (wobei in Frage gestellt werden kann, ob es sich bei dem technischen Spielzeug tatsächlich um einen Roboter handelt). Vor der Aufforderung, den Roboter mit dem Hammer zu zerschlagen, bekamen die Probanden entweder keine weiteren Informationen zum Roboter, eine personifizierte Geschichte zum Roboter (Name und Eigenschaften), oder eine Geschichte über Erfahrungen des Roboters, die dieser bereits gemacht hatte. Es zeigte sich, dass Probanden sich erheblich länger weigerten bzw. den Roboter gar nicht zerstörten, wenn der Roboter eine Hintergrundgeschichte besaß. Zwischen den beiden framing-Bedingungen gab es noch leichte Tendenzen zur „Erfahrungsgeschichte“ hinsichtlich des Einflusses auf das Verhalten der Probanden, dieser Unterschied wurde allerdings nicht signifikant.

Aufgrund dieser Ergebnisse wird die Hypothese aufgestellt, dass je reicher der Kontext ausgestaltet wird, desto stärker wird die (affektive) Bindung von Menschen gegenüber Robotern. Dies kann in bestimmten Kontexten wünschenswert sein, z. B. bei therapeutischen Robotereinsätzen (z. B. mit der Roboterrobbe Paro), im konventionellen Arbeitskontext ist Kontextmorphologie aber mit Vorsicht einzusetzen. Einerseits bietet framing ein hohes Potential um z. B. Berührungsängste abzubauen, andererseits ist eine affektive Bindung an Roboter aus ethischer Perspektive u. U. kritisch zu bewerten, da sich daraus ein gewisses unerwünschtes Manipulationspotential ergibt.

Der Einfluss von Kommunikations- als auch Kontextmorphologie sind abschließend in Tab. 4.6 zusammengefasst.

Tab. 4.6 Einfluss der Kommunikations- und Kontextmorphologie auf ethisch-soziale Aspekte der MRI

Beschreibungsmerkmal Morphologie des Roboters - Kommunikationsmorphologie			
Studie	UV-Variation	Entspricht in Taxonomie	Interpretation der Ergebnisse
Breazeal, Kidd, Thomaz, Hoffman, & Berlin, 2005	explizit non-verbal (EXP) vs. implizit + explizit non-verbal (EXP+IMP)	Kommunikation humanoid, weniger humanoid	Subjektiv: EXP+IMP: Probanden verstehen Handlungen des Roboters bzw. Roboter an sich besser & geben an, ein besseres mentales Modell des Roboters zu haben im Vgl. zu EXP Objektiv: EXP+IMP: Probanden brauchen weniger Zeit um Aufgabe zu erledigen, Roboter+Mensch machen weniger Fehler, Fehler des Roboters werden schneller detektiert <i>Hypothese:</i> Implizite non-verbale Kommunikation hat den höchsten dynamischen Informationsgehalt, erfordert in der Interpretation am wenigsten kognitive Ressourcen, ist kulturübergreifend eindeutig und ist daher immer zu bevorzugen. Kann bei Bedarf mit anderen Kommunikationsformen erweitert werden (z.B. bei geforderten Standards einer expliziten Bestätigung)
Beschreibungsmerkmal Morphologie des Roboters - Kontextmorphologie			
Studie	UV-Variation	Entspricht in Taxonomie	Interpretation der Ergebnisse
Darling u. a., 2015	zoomorph vs. funktional	Kontext zoomorph, funktional	Je reicher der Kontext, desto stärker die Bindung.

4.7 Zusammenfassung

Ziel der Analyse war, konkrete Gestaltungsmerkmale der Zusammenarbeit von Mensch und Roboter zu identifizieren, die ethisch-soziale Aspekte der MRI sowohl positiv als auch negativ beeinflussen könnten. Dazu wurden in einem top-down Verfahren insgesamt 14 empirische MRI-Studien ausgewählt. Die identifizierten Wirkgrößen wurden anhand der Taxonomie von Onnasch und Kollegen (2016) klassifiziert und die Taxonomie anhand der Ergebnisse weiter verfeinert. Die aktualisierte Taxonomie teilt die Merkmalsbeschreibung Morphologie weiter in die Dimensionen Kontextmorphologie, Gestaltmorphologie, Kommunikationsmorphologie und Bewegungsmorphologie auf.

Insgesamt wurden drei Einflussgrößen auf ethisch-soziale Aspekte der MRI durch die Analyse identifiziert: die Interaktionsrolle des Menschen, der Autonomiegrad des Roboters sowie die Morphologie des Roboters in Bezug zu Kontext, Gestalt, Kommunikation und Bewegung.

Für die Wahl der Interaktionsrolle gilt: Je stärker der Mensch in einer gemeinsamen Aufgabe die Interaktion führt, desto höher ist auch das Verantwortungsgefühl für die Gesamtaufgabe von Roboter und Mensch. Gleichzeitig akzeptieren Menschen Roboter im Arbeitskontext grundsätzlich in gleichgestellter als auch in führender Funktion.

Darüber hinaus weist auch der Autonomiegrad des Roboters einen wesentlichen Einfluss auf die wahrgenommene Verantwortung des Menschen für die Gesamtaufgabe auf. Positiv zu bewerten ist, dass eine hohe Roboterautonomie zu einer erhöhten Akzeptanz führt. Allerdings zeigen sich auch Hinweise auf zunehmende Manipulationsmöglichkeiten durch einen hohen Autonomiegrad.

Bezüglich der Morphologie weisen die Studien darauf hin, dass sowohl Gestalt- als auch Kontextmorphologie im Arbeitsrahmen relativ gering ausfallen sollten, da diese Aspekte die Wahrscheinlichkeit einer nichtintendierten affektiven Bindung stark erhöhen und somit auch das Manipulationspotential. Da nur eine Studie die Kommunikationsmorphologie behandelt, können hier nur hypothesengenerierende Schlüsse gezogen werden, welche jedoch einer experimentellen Validierung bedürfen. Ein positives Bild zeichnen die Befunde einer anthropomorphen Bewegungsgestaltung des Roboters. Dies führt zu einer verbesserten Koordinationsleistung zwischen Mensch und Roboter während sich keine Hinweise auf mögliche negative Konsequenzen zeigen.

5 Empfehlungen für eine ethisch-sozial angemessene MRI

Im Folgenden werden die Ergebnisse der Auswertung aus Kap. 4 dahingehend interpretiert, dass sich Empfehlungen für eine ethisch-sozial reflektierte Gestaltung der MRI ergeben. Die Empfehlungen betreffen die Ausgestaltung der Interaktionsrolle, des Autonomiegrades und verschiedene Ebenen der Morphologie des Roboters im industriellen MRI-Kontext.

Grundsätzlich können mit der Unterscheidung zwischen Autonomie, Interaktionsrolle, Gestaltmorphologie, Kommunikationsmorphologie, Bewegungsmorphologie und Kontextmorphologie verschiedene Einflussgrößen voneinander abgegrenzt werden, die sich jeweils spezifisch und unterschiedlich stark auf die identifizierten Problembereiche auswirken. In diesem Sinne erlaubt die Taxonomie nach Onnasch und Kollegen (2016) als analytisches Werkzeug eine zielgerichtete Interpretation empirischer Studien, aus der sich grundlegende Hinweise für eine systematische Gestaltung der MRI im Hinblick auf ethisch-soziale Entwicklungen ergeben.

Da die aktuelle Datenbasis noch sehr gering ausfällt, sollten neue Studien zur MRI kontinuierlich in die Analyse mit einfließen, so dass die Gestaltungsempfehlungen einer formativen Validierung unterzogen werden können. Letztendlich sollten die Empfehlungen darüber hinaus noch anhand systematischer Studien empirisch überprüft werden. Daher ist festzuhalten, dass eine gezielte und umfangreiche Erforschung von unterschweligen Wirkungen der MRI-Gestaltung auf ethisch-soziale Problembereiche insbesondere im industriellen Arbeitskontext noch aussteht.

5.1 Einfluss von Autonomiegrad und Interaktionsrolle auf Akzeptanz und Verantwortung

Zunächst kann auf der Grundlage der Überlegungen in Kap. 3 festgehalten werden, dass Begriffe wie „Führungsrolle“ oder „Teamleitung“ in Bezug auf die Interaktionsrolle von Robotern problematisch sind, da Roboter im Unterschied zu Menschen nicht über soziale Kompetenzen und Verantwortungsfähigkeit verfügen. Beide Eigenschaften sind aber eng mit den gängigen Konzepten von Führung oder Teamleitung verbunden. Entsprechende Charakterisierungen von Robotern können bestenfalls auf eine vermittelte „Führungsfunktion“ hinweisen, die tatsächlich aber einem (nicht unmittelbar an der Interaktion beteiligten) anderen Mitarbeiter zukommt. Angemessener ist es, in Bezug auf Roboter von einem technischen Steuerungsmittel der Interaktion bzw. der Arbeit zu sprechen, vergleichbar mit einem Taktgeber oder einer Handlungsvorgabe als Ergebnis informationsverarbeitender Prozesse. Eine vermenschliche Beschreibung des Roboters in der MRI kann dagegen zu Unklarheiten und unangemessenen Vorstellungen führen, indem beispielsweise eine ‚Ersetzung‘ der sozialen Kompetenz und der Verantwortungsdimension durch den Roboter impliziert wird, während diese Komponenten allerdings faktisch entfallen oder sich verschieben.

Voraussetzung für den Einsatz von autonomen Robotern in der MRI als Steuerungselemente der Arbeit sind darüber hinaus Akzeptanz und Vertrauen in die autonomen Fähigkeiten. In diesem Zusammenhang erscheint ein bestimmtes Maß an Transpa-

renz bezüglich der Funktionsweise und der Entscheidungsprinzipien empfehlenswert, um die Vorgaben für den menschlichen Mitarbeiter nachvollziehbar zu machen. Ein gewisses Maß an Transparenz kann darüber hinaus auch der Unklarheit über die Verantwortungszuordnung sowie einer suggestiven Beeinflussung des menschlichen Mitarbeiters gegen seine Interessen entgegenwirken. Die Nachvollziehbarkeit der ‚Entscheidungen‘ des Roboters sollte prinzipiell eine fundierte Einschätzung und Beurteilung der konkreten Situation durch den menschlichen Mitarbeiter zulassen, wodurch eine subjektive Sicherheit in der Meinungsbildung geschaffen wird.

Allerdings ist die Schaffung von Transparenz bezüglich der autonomen Funktionsweise des Roboters vornehmlich ein organisationaler Faktor und weniger ein Merkmal der direkten Interaktion, denn die konkrete Arbeitssituation in der MRI ist begrenzt durch die Komplexität und die Menge der sinnvoll aufzunehmenden Informationen – schließlich soll die Autonomie des Roboters den menschlichen Arbeiter in der MRI von Entscheidungen entlasten und eine Informationsüberflutung in komplexen Arbeitssituationen gerade verhindern.

Die Akzeptanz einer technischen Steuerung von Arbeitsprozessen hängt darüber hinaus auch davon ab, ob die damit erreichte Aufgabenallokation den jeweiligen Fähigkeiten und Kompetenzen der interagierenden Arbeiter angemessen ist. Hier erscheint es empfehlenswert, den menschlichen Arbeiter einen Beurteilungs- und Einflussspielraum einzuräumen, um einen Abgleich zwischen dem umfassenderen Überblick des Menschen über den Arbeitsprozess und der technischen Steuerung der MRI zu ermöglichen.

5.2 Einfluss der Morphologie auf Verantwortungsdiffusion und affektive Bindung

Bei der Formulierung von Empfehlungen zur Morphologie des Roboters in der MRI sind die verschiedenen Ebenen der äußeren Gestalt, der Bewegungsmorphologie, der Kommunikationsweise und der Kontextmorphologie zu unterscheiden.

Äußerliche humanoide oder zoomorphe Gestaltungsmerkmale eines Roboters sind im industriellen Arbeitskontext auf die jeweiligen Erfordernisse der gemeinsam auszuführenden Aufgabe abzustimmen und entsprechend zu minimieren. Entsprechende Merkmale können sich einerseits positiv auf die Kooperationsbereitschaft auswirken, andererseits sind Probleme der Verantwortungsdiffusion denkbar (vgl. Kap. 4.6.1). Darüber hinaus sind unangemessene emotionale Bindungen an den Roboter zu vermeiden.

Dagegen erscheint eine humanoide oder zoomorphe Bewegungsmorphologie in ethisch-sozialer Hinsicht unproblematisch und aufgrund ihrer generellen Vorteile in kooperativen und kollaborativen Interaktionen empfehlenswert. Sie kann die wahrgenommene Intentionalität des Roboters und folglich die Koordinationsleistung von Mensch und Roboter erhöhen. Es gibt keine Hinweise zu Auswirkungen der Bewegungsmorphologie auf die Entwicklung einer affektiven Bindung oder einer Verantwortungsdiffusion. Allerdings erscheint es nicht ratsam, eine mensch- oder tierähnliche Bewegungsgestaltung mit humanoider oder zoomorpher Gestalt- und Kontextmorphologie zu kombinieren, da sie die Effekte dieser Gestaltungsebenen verstärkt (vgl. Kap. 4.6.2).

Für die Kommunikation zwischen Mensch und Roboter liegt es nahe, non-verbale Kommunikation gegenüber sprachlicher Kommunikation zu bevorzugen, da Sprache humanoide Assoziationen begünstigt und so beispielsweise bestimmte Rollenbilder transportiert (männliche bzw. weibliche Stimme). Es kann darüber hinaus angenommen werden, dass sich implizite non-verbale Kommunikation im Vergleich zu expliziter non-verbaler bzw. verbaler Kommunikation günstig auf die Interaktion auswirkt, da sie als integraler Bestandteil der Aktion realisiert werden kann und somit keinen kognitiven Mehraufwand verursacht.

Beim Einsatz eines fiktiven Kontextes der MRI (Kontextmorphologie) ist ähnlich wie bei der Gestaltmorphologie zwischen den kooperationsfördernden und den affektiv bindenden Einflüssen abzuwägen. Es gilt: Je reicher der Kontext ausgestaltet wird, desto stärker wird die affektive Bindung des Menschen an den Roboter. Dies kann in bestimmten Kontexten wünschenswert sein, z. B. in therapeutischen Funktionen, im konventionellen Arbeitskontext sind entsprechende kontextbildende Maßnahmen aber nur mit Vorsicht und nicht in Kombination mit einer humanoiden oder zoomorphen Robotergestalt einzusetzen.

Zu berücksichtigen ist auch die Entwicklung der morphologischen Eigenschaften mit zunehmender Interaktionszeit: So wirken kontextbildende Maßnahmen zu Beginn der Interaktion stark, ihr Einfluss nimmt allerdings über die Zeit ab, da fiktive Kontexte keine Merkmale der Situation oder des Roboters selbst sind und somit zunehmend an Bedeutung verlieren (bzw. in Vergessenheit geraten). Im Unterschied zu diesem abnehmenden Effekt ist nicht zu erwarten, dass sich der Einfluss der Bewegungs- oder Kommunikationsmorphologie über die Zeit wesentlich ändert. Insbesondere eine erwartungskonforme humanoide oder zoomorphe äußere Gestalt könnte auf Dauer zu einer Personifizierung oder Individualisierung des Roboters führen, da eine gemeinsame „Historie“ geschaffen wird. Aus diesem Grund bleiben Auswirkungen wie affektive Bindungen der Gestaltmorphologie dauerhaft problematisch oder verstärken sich über die Zeit sogar.

5.3 Empfehlungen zur industriellen MRI-Gestaltung

Für eine übersichtliche Darstellung werden vorangegangene Empfehlungen und Klarstellungen in Abb. 5.1 verdichtend visualisiert. Die in der Abbildung dargestellten Illustrationen werden anschließend nochmals in Kurzform erläutert, der Reihenfolge von links oben nach rechts unten folgend.

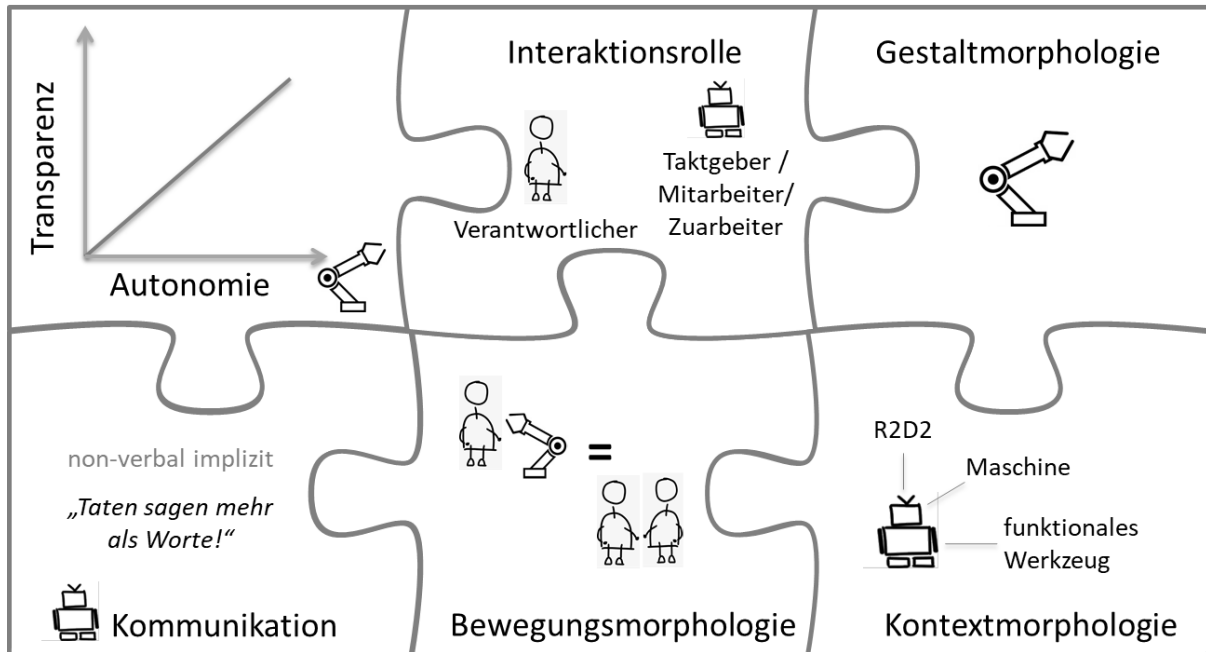


Abb. 5.1 Gestaltungsempfehlungen der industriellen MRI aus ethisch-sozialer Perspektive

Autonomie und Transparenz: Mit steigender Autonomie des Roboters sollte die Transparenz bzgl. seiner Funktionsweise und Entscheidungsgrundlagen ein wesentliches Gestaltungsmerkmal darstellen. Transparenz wirkt möglicher Verantwortungsdiffusion entgegen und kann eine Beeinflussbarkeit des Menschen abmindern.

Die Interaktionsrolle in Abhängigkeit der Aufgabenallokation: Grundsätzlich ist eine durch den Roboter geführte Interaktion unproblematisch wenn die Aufgabenallokation durch die Kompetenzabwägung zwischen Mensch und Roboter begründet ist. Der Roboter kann allerdings nur Führung im Sinne eines Taktgebers ausüben, der Mensch bleibt immer Verantwortlicher.

Empfehlungen zur Robotermorphologie

Gestaltmorphologie: Um Verantwortungsdiffusion und unangemessene emotionale Bindungen zu vermeiden, sollte eine anthropomorphe Robotergestaltung im Arbeitskontext immer relativ gering ausfallen.

Kommunikationsmorphologie: Um eine Übertragung menschlicher Stereotype zu vermeiden, ist eine implizite non-verbale Kommunikation zu bevorzugen.

Bewegungsmorphologie: Eine anthropomorphe Gestaltung ist grundsätzlich zu empfehlen, da dies die wahrgenommene Intentionalität und resultierend die Koordinationsleistung von Mensch und Roboter erhöht.

Kontextmorphologie: Im industriellen Anwendungsbereich sollte die Kontextmorphologie möglichst funktional gestaltet werden und sich an den technischen Merkmalen des Roboters orientieren.

6 Fazit: Abwägung der ethisch-sozialen Implikationen in der MRI

Das primäre Ziel des Einsatzes von Mensch-Roboter-Interaktionen im industriellen Arbeitskontext besteht schlussendlich darin, die Vorteile von menschlichen Arbeitern und robotischen Systemen möglichst optimal zu kombinieren. Roboter zeichnen sich durch Ausdauer, Präzision, Kraft, Schnelligkeit und Zuverlässigkeit aus, während Arbeiter flexibel, kreativ und verantwortungsbewusst sind. Unter der Voraussetzung technisch implementierter Sicherheitsstandards eröffnen Interaktionen von Menschen und Robotern die Aussicht auf eine grundsätzliche Verbesserung von Produktionsprozessen.

Im Rahmen der vorangegangenen Erörterung von ethisch-sozialen Aspekten der MRI ist deutlich geworden, dass sich das Ziel einer gelingenden und guten Zusammenarbeit zwischen Menschen und Robotern nicht alleine als technisches Problem stellt. Ein grundsätzlicher Ausgangspunkt der Überlegungen und Empfehlungen ist die Anerkennung einer prinzipiell sozialen Struktur von Mensch-Roboter-Interaktionen. In ethischer Hinsicht geht es folglich um die Untersuchung wünschenswerter Gestaltungen von MRI – und zwar nicht nur für den Bereich der ‚sozialen‘ Robotik im engeren Sinne, sondern auch für den industriellen Arbeitskontext, um den es hier geht. Denn auch wenn die technische Entwicklung zu einer sicheren und effizienten MRI in der industriellen Arbeit führt, so stellen sich Anschlussfragen an die Qualität des so geschaffenen Arbeitsumfelds und der sich darin entwickelnden Tätigkeiten aus menschlicher Perspektive.

In öffentlichen Debatten werden Fragen dieser Art im Vergleich zur Diskussion von Konsequenzen für Arbeitsmarkt und wirtschaftliche Produktivität nachrangig behandelt (Arlt et al., 2017). Allerdings verweisen beispielsweise Moniz & Krings (2016) auf die Notwendigkeit einer sozialwissenschaftlichen Erforschung von neuen Modellen der Arbeitsorganisation, die durch die MRI ermöglicht und nahegelegt werden. Entsprechende Forschungsfragen gehen über Probleme der technischen und wirtschaftlichen Sicherheit des Arbeiters und der Gesellschaft hinaus. Moniz & Krings schlagen in diesem Sinne vor, sozio-technische Konzepte für das jeweilige Arbeitsumfeld konkreter MRI-Szenarien zu erarbeiten. Das bedeutet, dass soziale Kategorien für die Interaktion von Robotern und Arbeitern entwickelt werden, mit dem Ziel, im Rahmen neuer Modelle der Arbeitsorganisation verbesserte Arbeitsbedingungen und eine höhere Arbeitsqualität zu realisieren. Relevant aus sozialwissenschaftlicher Perspektive sind in diesem Zusammenhang beispielsweise Lernprozesse, Kompetenzerwerb, dezentrale Entscheidungsfindung und partizipative Organisationsmodelle (Moniz & Krings, 2016).

In spezifisch ethischen Beurteilungen der sozialen Dimension bestimmter MRI-Szenarien geht es um eine kontrollierte Abwägung zwischen der Optimierung von Arbeitsabläufen und ethisch-sozial tendenziell problematischen Entwicklungen. So wird in den Empfehlungen, die in Kap. 5 auf der Basis der Literaturrecherche formuliert sind, die dringende Beachtung potenzieller Probleme wie Verantwortungsdiffusion, Instrumentalisierung und affektiver Bindung empfohlen. Mit diesen Begriffen sind allerdings keine fest umrissenen und einfach operationalisierbaren Probleme angesprochen; zunächst haben wir es mit sehr allgemeinen und potenziellen Szenarien

zu tun, in denen Werte berührt bzw. verletzt werden können, die nicht alleine auf Funktionalität und Effizienz zurückzuführen sind. Diese Identifikation der ethisch-sozialen Problembereiche und die daran orientierte Interpretation empirischer Studien verstehen sich als ein erster Vorschlag für eine Abschätzung der Technikfolgen in ethisch-sozialer Hinsicht und sind offen für weitere Argumentationen oder Kritik. Ob und wie einschneidend die entsprechenden Aspekte zu bewerten sind, muss Gegenstand weiterer Diskussionen und Untersuchungen sein. Auch die reale Gefahr für den Eintritt der identifizierten ethisch-sozialen Probleme wäre in der weiteren Forschung näher zu bestimmen – in der vorliegenden Arbeit können lediglich Tendenzen festgestellt werden, die im Vergleich zu empirisch bereits vorkommenden und sich abzeichnenden Anwendungsszenarien noch vergleichsweise abstrakt bleiben. Die Gestaltung der MRI in Richtung humanoider oder zoomorpher Roboter – und sei es auch nur auf der Ebene der Bewegungsmorphologie oder der Kontextualisierung – ist gelegentlich auch für die industrielle Arbeit ein gangbarer und in manchen Bereichen eingeschlagener Weg. Beispiele finden sich in den Robotern Baxter und Sawyer der Firma Rethink Robotics sowie dem workerbot der Firma pi4_robotics GmbH. Der workerbot wird darüber hinaus im Angebot der Firma Robozän GmbH gezielt als „Zeitarbeiter“ beschrieben und zum Mindestlohn zur Verfügung gestellt. Durch diesen Kontext werden gezielt Assoziationen zu menschlichen Arbeitern geweckt.

Wie in den vorhergehenden Abschnitten dargestellt, liegt ein sachlicher Grund für entsprechende Gestaltungen der MRI darin, dass Zuschreibungen intentionalen Verhaltens dadurch prinzipiell erleichtert werden, was wiederum den Interaktionen mit Robotern förderlich sein kann. Fähigkeiten und Kompetenzen, die vor dem Hintergrund der menschlichen Zusammenarbeit vorliegen und eingeübt sind, können so für die MRI nutzbar gemacht werden. Insgesamt kann ein Ziel der MRI-Gestaltung in diesem Sinne darin liegen, eine weitgehend intuitive Interaktion von Arbeitern und Robotern zu ermöglichen. Der Nutzen einer intuitiven Interaktion besteht nicht nur darin, dass unter Umständen der Aufwand für technische Schulungen reduziert und der Einsatz von Robotern im Arbeitskontext folglich flexibler und kostengünstiger erfolgen kann. Darüber hinaus kann intuitive MRI u.a. auch die Sicherheit erhöhen, hochkomplexe Arbeitsprozesse überschaubar machen und den Anteil menschlicher Arbeit insgesamt leichter erscheinen lassen (Moniz & Krings, 2016, S. 10).

Neben diesen direkten Auswirkungen auf den Arbeitsprozess sind psychologische Faktoren zu berücksichtigen, die den Einsatz von kooperierenden und kollaborierenden Robotern vor dem Hintergrund der bisherigen technischen Entwicklung von Industrierobotern unterstützen. Industrielle Arbeitsplätze waren in der Vergangenheit ausschließlich und sind auch heute noch überwiegend durch Schutzzäune räumlich von den lebensgefährlichen Industrierobotern getrennt. Das Bild des Industrieroboters war insofern bis zum Aufkommen der kooperativen und kollaborierenden Roboter aus der Perspektive des Arbeiters grundsätzlich mit Gefahr verbunden. Vor diesem industriehistorischen Hintergrund stellt sich nicht nur die technische Herausforderung, einen sicheren und dennoch leistungsfähigen Einsatz robotischer Technologien zu entwickeln, sondern auch den Arbeiter auf die neue Situation einzustellen. Ein Roboter sollte nicht bedrohlich wirken, um (im wörtlichen Sinne) Berührungsängste bei der Einführung der neuen Technologie abzubauen. Hier bieten sich nicht zuletzt auch ‚freundliche‘ Gestaltungen von Robotern im Sinne von humanoiden oder zoomorphen Gestaltungen an. In manchen Fällen wirkt bereits die Namensgebung in

diese Richtung, wie beispielsweise an den Modellen Baxter und Sawyer deutlich wird.

Intuitiv gestaltete Interaktionen und der Abbau von Berührungängsten sind zwei Beispiele für erwünschte Eigenschaften der MRI, die gegenüber den ethisch-moralisch problematischen Tendenzen der Vermenschlichung von Robotern abzuwägen sind. In Bezug auf die Gestaltung dieser Eigenschaften ist eine transparente Gegenüberstellung von Zielen und Problemen anzustreben, auf deren Grundlage eine bewusste Entscheidung zugunsten einer bestimmten Tendenz ermöglicht wird. Eine ähnliche Situation liegt mit der Eigenschaft der Autonomie von Robotern vor. Einerseits werden dem Arbeiter in der Interaktion mit hochgradig autonomen Robotern überfordernde, unnötige oder unangemessene Tätigkeiten abgenommen. Andererseits geht damit unter Umständen ein gemindert Verantwortungsbewusstsein einher, nicht zuletzt dann, wenn die Tätigkeiten des autonomen Roboters in bestimmten Fällen nicht sofort oder eindeutig erkennbar sind. Entsprechende ‚unkritische‘ Interaktionen können beispielsweise durch die Förderung von Transparenz über die ‚Entscheidungsprinzipien‘ des Roboters verhindert werden. Damit wird aber zugleich die entlastende Funktion der Autonomie des Roboters wiederum etwas eingeschränkt, oder es wird ein gewisser Schulungsaufwand erforderlich. Es gilt also Abzuwägen zwischen den positiven Effekten autonomer Robotik und deren Kehrseite, z. B. der Verantwortungsdiffusion.

Eine allgemeinere Tendenz zur ‚Vermenschlichung‘ von Robotern lässt sich nicht zuletzt an der Begrifflichkeit beobachten, mit der über die MRI auch im industriellen Arbeitskontext gesprochen wird. Die bereits vielfach erwähnten Metaphern des ‚Kollegen Roboter‘ und des ‚hybriden Teams‘ sind nur zwei Beispiele für eine suggestive Übertragung von menschlichen Eigenschaften auf technische Systeme, die sicherlich nicht folgenlos bleibt. So können beispielsweise Bilder aus der Science-Fiction transportiert werden, etwa um den Eindruck revolutionärer Innovation zu erzeugen. Oder es wird Akzeptanz nahegelegt, indem hochtechnologischen Entwicklungen ein ‚Gesicht‘ gegeben wird – möglicherweise gar im wörtlichen Sinne. Insgesamt entstehen Eindrücke der MRI, in denen eine (weitere) Entgrenzung zwischen Mensch und Technik anklingt (Kehl & Coenen, 2016).

Auch wenn die Übertragung menschlicher Eigenschaften auf sachliche Gründe zurückgehen mag – insbesondere auf die Zielvorstellungen von technischer Autonomie und intentional gerichteter Interaktion – kann sie im Vergleich zur Debatte über andere Hochtechnologien durchaus unangemessen wirken. Die in Kap. 5 formulierten Empfehlungen greifen in diesem Zusammenhang insbesondere das Phänomen einer affektiven Bindung zwischen Arbeiter und Roboter auf. Es wäre darüber hinaus zu untersuchen, welche Konsequenzen im Vergleich dazu eine gezielte ‚Verfremdung‘ des Roboters auf die MRI haben könnte, d. h. eine bewusst nicht-humanoid bzw. nicht-zoomorphe Gestaltung. Beispielsweise könnte in bestimmten Szenarien ein gegen die Gewohnheit absichtlich technisch gehaltener Kontext eine affektive Bindung zum Roboter verhindern oder zumindest relativieren. Auch in Szenarien, in denen Roboter in der Cloud nach außen unsichtbar vernetzt sind, woraus eine fließende Anpassung und Veränderung des konkreten Roboter-Gegenübers in der MRI resultieren könnte, wird der Eindruck eines quasi-humanoiden ‚Individuums‘ offenbar gebrochen. In dieser Richtung stellen sich interessante Anschlussfragen.

Schließlich gibt es auch Hinweise auf eine Entwicklung der MRI, in der der Roboter nicht so sehr als unabhängig agierendes Gegenüber aufgefasst wird, sondern vielmehr als technische Erweiterung des handelnden Menschen (Carpenter, 2016). Hier wäre insofern nicht mehr vom ‚Kollegen Roboter‘ oder vom ‚hybriden Team‘ zu sprechen, sondern vielmehr von einem besonders komplexen und weitreichenden Werkzeug, dass – wenn auch nicht unmittelbar räumlich – als Verlängerung des eigenen Handelns empfunden wird. Diese Idee einer Mensch-Roboter-Interaktion könnte zu einer grundsätzlich andersartigen Begrifflichkeit führen, in der sich die Beziehung zwischen Mensch und Roboter anders darstellen würde. In dieser Richtung ergeben sich Forschungsfragen zur begrifflichen Bestimmung und Gestaltung der MRI und zur Debatte über ihre weitere Entwicklung, die für eine ethisch-soziale Beurteilung relevant sein dürften.

Diese und weitere Anschlussfragen an die künftige Forschung können im Rahmen des vorliegenden Berichts, in dem die ethischen und sozialen Implikationen der Mensch-Roboter-Interaktion thematisiert wurden, nur angedeutet werden. Die in den Kapiteln 2 und 3 angestellten Überlegungen sowie die auf Basis der Literaturrecherche in Kap. 4 und 5 formulierten Empfehlungen zu einer ethisch-sozialen Gestaltung der MRI betreffen einige der spezifischen Probleme einer zunehmenden Autonomie technischer Systeme sowie einer impliziten Vermenschlichung von Technik in Morphologie und Begrifflichkeit. Ihre Untersuchung und die Einschätzung ihrer Auswirkungen im Kontext der industriellen Produktion bilden erste Schritte für die Einbeziehung von Aspekten, die wesentlich zur positiven Entwicklung der Arbeit in der Industrie 4.0 beitragen können.

Literaturverzeichnis

- Allen, C., Varner, G., & Zinser, J. (2000). Prolegomena to any future artificial moral agent. *Journal of Experimental & Theoretical Artificial Intelligence*, 12(3), 251–261. <https://doi.org/10.1080/09528130050111428>
- Anderson, M., & Anderson, S. L. (2011). *Machine ethics*. New York: Cambridge University Press.
- Applin, S. A., & Fischer, M. D. (2015). Cooperation Between Humans and Robots: Applied Agency in Autonomous Processes. In L. D. Riek, W. Hartzog, D. Howard, A. Moon, & R. Calo (Eds.), *10th ACM/IEEE International Conference on Human/Robot Interaction, The Emerging Policy and Ethics of Human Robot Interaction workshop*. Portland (Oregon). Retrieved from http://www.openroboethics.org/hri15/wp-content/uploads/2015/02/Af-Applin_Fischer.pdf
- Arlt, H.-J., Kempe, M., & Osterberg, S. (2017). *Die Zukunft der Arbeit als öffentliches Thema. Presseberichterstattung zwischen Mainstream und blinden Flecken*. Frankfurt am Main: Otto Brenner Stiftung.
- Arnold, T., & Scheutz, M. (2016). Against the moral Turing test: accountable design and the moral reasoning of autonomous systems. *Ethics and Information Technology*, 18(2), 103–115. <https://doi.org/10.1007/s10676-016-9389-x>
- Asaro, P. M. (2006). What Should We Want From a Robot Ethic? *International Review of Information Ethics*, 6, 9–16.
- Atmaca, S., Sebanz, N., Prinz, W., & Knoblich, G. (2008). Action co-representation: The joint SNARC effect. *Social Neuroscience*, 3(3-4), 410–420. <https://doi.org/10.1080/17470910801900908>
- Bauer, A., Wollherr, D., & Buss, M. (2008). Human-Robot Collaboration: A Survey. *International Journal of Humanoid Robotics (IJHR)*, 5(4), 47–66.
- Beck, S. (2009). Grundlegende Fragen zum rechtlichen Umgang mit der Robotik. *Juristische Rundschau*, (6).
- Birnbacher, D. (2013). Ethik und Robotik - Wie weit trägt die Analogie der Tierethik? In E. Hilgendorf & J.-P. Günther (Eds.), *Robotik und Gesetzgebung (Vol. 2, pp. 303–319)*. Baden-Baden: Nomos.
- Birnbacher, D., & Birnbacher, W. (2016). *Automatisiertes Fahren*. *Information Philosophie*.
- Breazeal, C., Kidd, C. D., Thomaz, A. L., Hoffman, G., & Berlin, M. (2005). Effects of nonverbal communication on efficiency and robustness in human-robot teamwork. In *Intelligent Robots and Systems, 2005.(IROS 2005). 2005 IEEE/RSJ International Conference on (pp. 708–713)*. IEEE. Retrieved from <http://ieeexplore.ieee.org/abstract/document/1545011/>

Briggs, G., Gessell, B., Dunlap, M., & Scheutz, M. (2014). Actions speak louder than looks: Does robot appearance affect human reactions to robot protest and distress? In *Robot and Human Interactive Communication, 2014 RO-MAN: The 23rd IEEE International Symposium on* (pp. 1122–1127). IEEE. Retrieved from <http://ieeexplore.ieee.org/abstract/document/6926402/>

Briggs, G., & Scheutz, M. (2014). How robots can affect human behavior: Investigating the effects of robotic displays of protest and distress. *International Journal of Social Robotics*, 6(3), 343–355.

Campa, R. (2015). *Humans and automata: a social study of robotics*. Frankfurt am Main; Bern: Peter Lang.

Carpenter, J. (2016). *Culture and human-robot interaction in militarized spaces: a war story*. Burlington, VT: Ashgate Publishing Company.

ChrSTALLER, T., Decker, M., GILSBACH, J.-M., Hirzinger, G., Lauterbach, K., Schweighofer, E., Sturma, D. (2001). *Robotik - Perspektiven für menschliches Handeln in der zukünftigen Gesellschaft*. Berlin: Springer.

Darling, K. (2012). Extending legal protection to social robots: The effects of anthropomorphism, empathy, and violent behavior towards robotic objects. Retrieved from https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2044797

Darling, K. (2015). “Who’s Johnny?” Anthropomorphic Framing in Human-Robot Interaction, Integration, and Policy. *SSRN Electronic Journal*. <https://doi.org/10.2139/ssrn.2588669>

Darling, K., Nandy, P., & Breazeal, C. (2015). Empathic concern and the effect of stories in human-robot interaction. In *24th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*. IEEE. Retrieved from <http://ieeexplore.ieee.org/abstract/document/7333675/>

Datteri, E., & Tamburrini, G. (2013). Robotic Weapons and Democratic Decision-Making. In E. Hilgendorf & J.-P. Günther (Eds.), *Robotik und Gesetzgebung* (Vol. 2, pp. 211–229). Baden-Baden: Nomos.

Decker, M. (2013). Robotik. In A. Grunwald (Ed.), *Handbuch Technikethik* (pp. 354–358). Stuttgart: J.B. Metzler.

Floridi, L., & Sanders, J. W. (2004). On the Morality of Artificial Agents. *Minds and Machine*, 14, 349–379.

Goetz, J., Kiesler, S., & Powers, A. (2003). Matching robot appearance and behavior to tasks to improve human-robot cooperation. In *Robot and Human Interactive Communication, 2003. Proceedings. ROMAN 2003. The 12th IEEE International Workshop on* (pp. 55–60). IEEE. Retrieved from <http://ieeexplore.ieee.org/abstract/document/1251796/>

- Gombolay, M. C., Gutierrez, R. A., Clarke, S. G., Sturla, G. F., & Shah, J. A. (2015). Decision-making authority, team efficiency and human worker satisfaction in mixed human–robot teams. *Autonomous Robots*, 39(3), 293–312. <https://doi.org/10.1007/s10514-015-9457-9>
- Gunkel, D. J. (2012). *The machine question: critical perspectives on AI, robots, and ethics*. Cambridge (Massachusetts); London: MIT Press.
- Hilgendorf, E. (2012). Können Roboter schuldhaft handeln? Zur Übertragbarkeit unseres normativen Grundvokabulars auf Maschinen. In S. Beck (Ed.), *Jenseits von Mensch und Maschine. Ethische und rechtliche Fragen zum Umgang mit Robotern, Künstlicher Intelligenz und Cyborgs* (Vol. 1, pp. 119–132). Baden-Baden: Nomos.
- Hindriks, F. (2011). Control, intentional action, and moral responsibility. *Philosophical Psychology*, 24(6), 787–801. <https://doi.org/10.1080/09515089.2011.562647>
- Hinds, P. J., Roberts, T. L., & Jones, H. (2004). Whose job is it anyway? A study of human-robot interaction in a collaborative task. *Human-Computer Interaction*, 19(1), 151–181.
- Hooper, P., Spencer, M., & Slok, T. (2016). Using the Fed’s model to evaluate Mr. Trump’s fiscal proposals. *Global Economic Perspectives*. Deutsche Bank Research.
- Jonas, H. (2003). *Das Prinzip Verantwortung: Versuch einer Ethik für die technologische Zivilisation* (1. Auflage). Frankfurt am Main: Suhrkamp.
- Kahn Jr, P. H., Kanda, T., Ishiguro, H., Gill, B. T., Ruckert, J. H., Shen, S., ... Severson, R. L. (2012). Do people hold a humanoid robot morally accountable for the harm it causes? In *Proceedings of the seventh annual ACM/IEEE international conference on Human-Robot Interaction* (pp. 33–40). ACM. Retrieved from <http://dl.acm.org/citation.cfm?id=2157696>
- Kehl, C., & Coenen, C. (2016). *Technologien und Visionen der Mensch-Maschine-Entgrenzung* (Vol. 167, p. S. 170). Berlin: Büro für Technikfolgen-Abschätzung beim Deutschen Bundestag (TAB).
- Kim, T., & Hinds, P. (2006). Who should I blame? Effects of autonomy and transparency on attributions in human-robot interaction. In *Robot and Human Interactive Communication, 2006. ROMAN 2006. The 15th IEEE International Symposium on* (pp. 80–85). IEEE. Retrieved from <http://ieeexplore.ieee.org/abstract/document/4107789/>
- Krüger, J., Lien, T. K., & Verl, A. (2009). Cooperation of human and machines in assembly lines. *CIRP Annals - Manufacturing Technology*, 58(2), 628–646. <https://doi.org/10.1016/j.cirp.2009.09.009>
- Kuz, S., Mayer, M. P., Müller, S., & Schlick, C. M. (2013). Using anthropomorphism to improve the human-machine interaction in industrial environments (part i). In *International Conference on Digital Human Modeling and Applications in Health, Safety, Ergonomics and Risk Management* (pp. 76–85). Springer. Retrieved from http://link.springer.com/chapter/10.1007/978-3-642-39182-8_9

Marino, D., & Tamburrini, G. (2006). Learning robots and human responsibility. *International Review of Information Ethics*, 6(12), 46–51.

Mayer, M. P., Kuz, S., & Schlick, C. M. (2013). Using anthropomorphism to improve the human-machine interaction in industrial environments (part II). In *International Conference on Digital Human Modeling and Applications in Health, Safety, Ergonomics and Risk Management* (pp. 93–100). Springer. Retrieved from http://link.springer.com/chapter/10.1007/978-3-642-39182-8_11

Moniz, A. B., & Krings, B.-J. (2016). Robots Working with Humans or Humans Working with Robots? Searching for Social Dimensions in New Human-Robot Interaction in Industry. *Societies*, 6(3), 23.

Mudry, P.-A., Degallier, S., & Billard, A. (2008). On the influence of symbols and myths in the responsibility ascription problem in roboethics - A roboticist's perspective. In *RO-MAN*.

Neuhäuser, C. (2014). Roboter und moralische Verantwortung. In E. Hilgendorf (Ed.), *Robotik im Kontext von Recht und Moral* (Vol. 3, pp. 269–286). Baden-Baden: Nomos.

Onnasch, L., Maier, X., & Jürgensohn, T. (2016). Mensch-Roboter-Interaktion-Eine Taxonomie für alle Anwendungsfälle. *Baua: Fokus, Bundesanstalt Für Arbeitsschutz Und Arbeitsmedizin*, 1–12. <https://doi.org/10.21934/baua:fokus20160630>

Parise, S., Kiesler, S., Sproull, L., & Waters, K. (1996). My partner is a real dog: Cooperation with social agents. In *Proceedings of the 1996 ACM conference on Computer supported cooperative work* (pp. 399–408). ACM. Retrieved from <http://dl.acm.org/citation.cfm?id=240351>

Pereira, L. M., & Saptawijaya, A. (2016). *Programming machine ethics*. Berlin, Heidelberg: Springer.

Rammert, W., & Schulz-Schaeffer, I. (2002). *Können Maschinen handeln? Soziologische Beiträge zum Verhältnis von Mensch und Technik*. Frankfurt am Main: Campus.

Richards, N. M., & Smart, W. D. (2016). How Should the Law Think About Robots? In R. Calo, Froomkin, M., & Kerr, I. (Eds.), *Robot Law* (pp. 3–22). Northampton: Edward Elgar.

Riek, L. D., Rabinowitch, T.-C., Bremner, P., Pipe, A. G., Fraser, M., & Robinson, P. (2010). Cooperative gestures: Effective signaling for humanoid robots (pp. 61–68). IEEE. <https://doi.org/10.1109/HRI.2010.5453266>

Riek, L. D., Rabinowitch, T.-C., Chakrabarti, B., & Robinson, P. (2009). Empathizing with robots: Fellow feeling along the anthropomorphic spectrum. In *Affective Computing and Intelligent Interaction and Workshops, 2009. ACII 2009. 3rd International Conference on* (pp. 1–6). IEEE. Retrieved from <http://ieeexplore.ieee.org/abstract/document/5349423/>

Rizzolatti, G., & Craighero, L. (2004). The Mirror-Neuron System. *Annual Review of Neuroscience*, 27(1), 169–192.
<https://doi.org/10.1146/annurev.neuro.27.070203.144230>

Scholtz, J. (2002). Human-robot interactions: Creating synergistic cyber forces. In *System Sciences, 2003. Proceedings of the 36th Annual Hawaii International Conference on* (p. 10–pp). IEEE. Retrieved from
http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=1174284

Sebanz, N., & Knoblich, G. (2009). Prediction in Joint Action: What, When, and Where. *Topics in Cognitive Science*, 1, 353–367.

Sebanz, N., Knoblich, G., & Prinz, W. (2003). Representing others' actions: just like one's own? *Cognition*, 88(3), B11–B21.

Sebanz, N., Knoblich, G., & Prinz, W. (2005). How Two Share a Task: Corepresenting Stimulus-Response Mappings. *Journal of Experimental Psychology: Human Perception and Performance*, 31(6), 1234–1246. <https://doi.org/10.1037/0096-1523.31.6.1234>

Stubbs, K., Hinds, P. J., & Wettergreen, D. (2007). Autonomy and Common Ground in Human-Robot Interaction: A Field Study. *IEEE Intelligent Systems*, 22(2), 42–50.
<https://doi.org/10.1109/MIS.2007.21>

Sullins, J. P. (2006). When Is a Robot a Moral Agent? *International Review of Information Ethics*, 6, 23–30.

Tzafestas, S. G. (2016). *Roboethics - A Navigating Overview*. New York, Berlin, Heidelberg: Springer.

Verein Deutscher Ingenieure, A. G. der T. (1993). VDI Richtlinie 3780. In H. Lenk & G. Ropohl (Eds.), *Technik und Ethik* (pp. 334–364). Stuttgart: Reclam.

Veruggio, G. (2005). The birth of roboethics. Retrieved from
<https://philpapers.org/rec/VERTBO-3>

Veruggio, G., & Operto, F. (2008). Roboethics - Social and Ethical Implications of Robotics. In B. Siciliano & O. Khatib (Eds.), *Springer Handbook of Robotics* (pp. 1499–1524). Berlin, Heidelberg: Springer.

Wagner, K., Rex, B. F., & Eicher, M. (2003). *Praktische Personalführung*. Wiesbaden: Gabler Verlag. Retrieved from <http://link.springer.com/10.1007/978-3-322-88994-2>

Wallach, W., & Allen, C. (2009). *Moral machines: teaching robots right from wrong*. Oxford; New York: Oxford University Press.

Yanco, H. A., & Drury, J. L. (2002). A taxonomy for human-robot interaction. In Proceedings of the AAAI Fall Symposium on Human-Robot Interaction (pp. 111–119). Retrieved from <http://www.aaai.org/Papers/Symposia/Fall/2002/FS-02-03/FS02-03-015.pdf>

Yanco, H. A., & Drury, J. L. (2004). Classifying human-robot interaction: an updated taxonomy. In SMC (3) (pp. 2841–2846). Retrieved from [http://www.eecs.tufts.edu/~mpoor01/DiscertationStuff/Animatronics%20\[Robotics\]/yanco-drury-tax-smc04.pdf](http://www.eecs.tufts.edu/~mpoor01/DiscertationStuff/Animatronics%20[Robotics]/yanco-drury-tax-smc04.pdf)

Abbildungsverzeichnis

Abb. 4.1	Schematische Darstellung der MRI-Taxonomie: Klassifikationscluster und Beschreibungsmerkmale	45
Abb. 4.2	Canvas-Darstellung der MRI-Taxonomie	46
Abb. 4.3	Canvas-Darstellung der MRI-Taxonomie mit Erweiterung der Dimension "Morphologie des Roboters"	49
Abb. 4.4	Versuchsaufbau der Studie von Breazeal und Kollegen (2005)	54
Abb. 5.1	Gestaltungsempfehlungen der industriellen MRI aus ethisch-sozialer Perspektive	61

Tabellenverzeichnis

Tab. 4.1	In die Analyse eingegangenen Studien, unterteilt nach adressierten ethischen Aspekten	40
Tab. 4.2	Einfluss der Interaktionsrolle des Menschen auf ethisch-soziale Aspekte der MRI	50
Tab. 4.3	Einfluss der Roboterautonomie auf ethisch-soziale Aspekte der MRI	51
Tab. 4.4	Einfluss der Gestaltmorphologie auf ethisch-soziale Aspekte der MRI	52
Tab. 4.5	Einfluss der Bewegungsmorphologie auf ethisch-soziale Aspekte der MRI	53
Tab. 4.6	Einfluss der Kommunikations- und Kontextmorphologie auf ethisch-soziale Aspekte der MRI	56