**Veröffentlicht in** „Review of Philosophy“.
and Psychology **11, Seiten 881–897**
**(2020)“, die als Verweis auf diese Arbeit**
**zitiert werden sollten.**

# KI verstehen – können und sollten wir uns einfühlen mit Robotern?

Susanne Schmetkamp1

Published online: 28 April 2020

**Abstrakt**

**Dieser Artikel weitet die Debatte über Empathie mit Menschen, Tieren oder fiktiven Charakteren auf Mensch-Roboter-Beziehungen aus und schlägt zwei verschiedene Perspektiven vor, um den Umfang und die Grenzen der Empathie mit Robotern zu beurteilen: Die erste ist erkenntnistheoretischer Natur, während die zweite normativ ist. Dabei hilft uns der erkenntnistheoretische Ansatz um zu klären, ob wir uns in künstliche Intelligenz hineinversetzen können, oder genauer: in soziale Roboter. Das Haupträtsel hier besteht unter anderem darin, was genau es ist mit denen wir uns identifizieren können, wenn Roboter keine Emotionen oder Überzeugungen haben, da sie keine haben ein Bewusstsein in einem ausführlicheren Sinne. Allerdings durch den Vergleich von Robotern mit fiktiven Charaktere zeigt die Arbeit, dass wir uns immer noch in Roboter hineinversetzen können und dass viele von ihnen Bestehende Berichte über Empathie und Gedankenlesen sind mit einer solchen Ansicht vereinbar. Damit Dabei konzentriert sich das Papier auf die Bedeutung der Perspektivenübernahme und der Behauptungen, die wir aufstellen schreiben Robotern auch so etwas wie ein perspektivisches Erlebnis zu. Der normative Ansatz untersucht die moralischen Auswirkungen der Empathie mit Robotern. In diesem Zusammenhang das Papier Erörtert kritisch drei mögliche Antworten: strategisch, antibarbarisierend und pragmatisch. Die letztgenannte Position wird verteidigt, indem betont wird, dass wir zunehmend gezwungen sind mit Robotern in einer gemeinsamen Welt zu interagieren und Roboter in unsere Moral aufzunehmen Rücksichtnahme sollte als integraler Bestandteil unseres Selbst- und Fremdverständnisses angesehen werden.**

**Schlüsselwörter Empathie. künstliche Intelligenz. Humanoide Roboter. Interaktion. Perspektiven einnehmen. Fiktive Charaktere. Ethik**

## 1. Einleitung

**Debatten über Empathie oder allgemeiner zwischenmenschliches Verständnis waren eine Hauptstütze der Wissenschaft in einem breiten Spektrum von Disziplinen. Allerdings hat es zwar viel gegeben wurde über die menschliche Fähigkeit geschrieben, sich in reale Menschen oder fiktive Charaktere hineinzuversetzen**

*** Susanne Schmetkamp**
**Susanne.schmetkamp@icloud.com**

1 **Institut für Philosophie, Universität Freiburg, Freiburg, Schweiz**

(für aktuelle Übersichten siehe Coplan und Goldie 2011; Maibom 2017) haben Philosophen bis vor kurzem die Rolle der Empathie in Mensch-Roboter-Interaktionen (HRI) etwas vernachlässigt (vgl. Brinck und Balkenius 2018; Lin et al. 2017). Doch im Einklang mit der wachsenden Zahl von Studien zu Emotionen oder anderen Merkmalen künstlicher Intelligenzsysteme1 besteht ein großes philosophisches Interesse an der Möglichkeit und Notwendigkeit der Interaktion und Empathie mit verschiedenen Formen künstlicher Intelligenz, insbesondere mit sogenannten sozialen Roboter.2 Dieses Interesse hat auch zu Diskussionen über den Wert von Empathie für die Gesellschaft im Allgemeinen oder für die Gesundheitsversorgung und Therapie im Besonderen geführt (Coeckelbergh 2018; Darling 2016; Engelen 2018; Loh 2019; Misselhorn im Druck; Vallor 2011). Es zeichnet sich ab, dass Roboter und Androiden – also Roboter, die wie Menschen aussehen – in Zukunft zu mehr oder weniger eigenständigen Akteuren mit sozialen Fähigkeiten werden. Als solche werden sie zu wichtigen Begleitern und zunehmend in der Lage, Beziehungen zu Menschen aufzubauen (Benford und Malartre 2007; Breazeal 2002; Dumouchel und Damiano 2017). Darüber hinaus werden Deep-Learning-Systeme (Kasparov 2017) in vielen (bislang) menschlichen Berufen zum Einsatz kommen, was nicht nur einige Aufgaben oder Herausforderungen (z. B. in der medizinischen Forschung) verbessern oder erleichtern wird; Sie könnten uns auch dazu zwingen, einige Schlüsselkonzepte wie Intelligenz, Entscheidungsfreihei

Wie Studien gezeigt haben (Leite et al. 2013), hängen Form und Erfolg von Mensch-Roboter-Beziehungen oft von menschenähnlichen Eigenschaften ab – etwa der Fähigkeit der Roboter, Emotionen auszudrücken, zu interagieren und (mehr oder weniger) autonome Entscheidunge Diese Fähigkeiten sind auch wichtig für das gegenseitige empathische Verstehen.3 Während der Mensch Emotionen auch in Bezug auf abstrakte virtuelle Formen oder sogar in Bezug auf technische Geräte (beste Beispiele sind Smartphones und Computer) erkennt und zuschreibt, für unsere kooperative und kollaborative Interaktion mit Robotern – insbesondere im medizinischen oder gesundheitsbezogenen Kontext – könnte eine starke menschliche Ähnlichkeit für den Erfolg dieser Interaktionen entscheidend sein. Mit der zunehmenden Präsenz humanoider sozialer Roboter in der Gesellschaft steigt auch die Notwendigkeit, unsere Interaktionen mit ihnen zu untersuchen und zu gestalten. Die aktuelle Robotergeneration ist bereits in der Lage, vielfältige Emotionen auszudrücken – die humanoide KI „Sophia“ beispielsweise kennt 60 verschiedene Gesichtsausdrücke und scheint sogar in der Lage zu sein, mit Humor und Ironie zu kommunizieren. Allerdings verfügen Roboter nicht über ein Bewusstsein im Sinne subjektiver Erfahrung4 und sie besitzen weder Humor noch Emotionen im ausgefeilten Sinne (Boden 2016; MacLennan 2014; Scheutz 2011). Dennoch könnten sie etwas haben, das als Analogie zu menschlichen Emotionen und einigen mentalen Prozessen angesehen werden kann. Darüber hinaus könnte es angesichts neuerer Erkenntnisse aus der Philosophie der verkörperten Kognition der Fall sein, dass der menschenähnliche Körper und das Verhalten sowie die „erweiterte“ Kognition den

[1] Eine Gruppe am MIT Media Lab und der IEEE Standards Association plädiert für das Konzept der „erweiterten“ Intelligenz statt der „künstlichen“. Mit einem solchen neuen Narrativ des „Erweiterten“ wollen sie sicherstellen, dass Roboter den Menschen nicht unterstützen, sondern mit ihm kooperieren. Gemeinsam gründeten sie den Council on Extended Intelligence CXI, siehe https://globalcxi.org (letzter Zugriff 12.12.2019).

[2] Ein vom ERC finanziertes Projekt an der Universität Glasgow unter der Leitung von Emily Cross untersucht insbesondere die Sozialisierung von Menschen mit künstlicher Intelligenz und die Bedeutung der Interaktion und Beziehungen mit Robotern für die soziale Kognition. Ein Schwerpunkt liegt auf der Fähigkeit von Robotern, Begleiter zu sein, http://www.so-bots.com (letzter Zugriff 20.12.2019).

[3] Zum Phänomen des „Uncanny Valley“ siehe unten.

[4] Zumindest wenn wir eine antiphysikalische Position vertreten.

sie in anderen (Benford und Malartre 2007, 181; Hoffmann und Pfeifer 2018; Newen et al. 2018). 5

Empathie wird allgemein als eine entscheidende Möglichkeit angesehen, die mentalen Zustände anderer durch Gedankenlesen, emotionales Teilen und/oder oder oder erfahrungsmäßiges Miterleben (siehe z. B. Engelen und Röttger-Rössler 2012; Goldman 2006; Stueber 2018; Zahavi 2014). 6 In der Philosophie wird Empathie üblicherweise von affektiver Ansteckung und moralischer Sympathie bzw. Mitgefühl unterschieden. 7 Wobei letztere auf das Wohl abzielt Empathie führt in erster Linie dazu, die mentalen Prozesse anderer zu verstehen – etwa Emotionen oder Überzeugungen . Im Gegensatz zur bloßen emotionalen Ansteckung muss eine Selbst-Andere-Differenzierung vorhanden sein (De Vignemont und Jacob 2012). Zu diesem Punkt gibt es weiterhin erhebliche Debatten, und es wurden verschiedene Definitionen und Ansätze vorgeschlagen, die sich mit Fragen befassen wie: Wie nehmen wir die Zustände und Erfahrungen anderer wahr und wie nehmen wir darauf zu? Wie ist der empathische P Was ist das Ergebnis dieses Prozesses? Im Großen und Ganzen sind die vorherrschenden Theorien – aus der Philosophie des Geistes oder der Phänomenologie – die Spiegelneuronen- oder Resonanztheorie (MNT) (Gallese 2001), die Theorietheorie (TT) (Fodor 1987; Gopnik und Wellman 1994), die Simulationstheorie (ST ) (De Vignemont und Jacob 2012; Goldman 2006, 2011; Stueber 2006), Direct Perception Theory (DPT) (Zahavi 2011) mit ihren Variationen von Interaction Theory (IT) (Gallagher 2008, 2017) und Narrativity Theory (NT) (Gallagher und Hutto 2008). Darüber hinaus gibt es hybride und pluralistische Theorien, die zwei oder mehr Ansätze kombinieren, etwa direkte Wahrnehmung und Imagination8 (Schmetkamp 2017, 2019; Dullstein 2013; für hervorragende Übersichten siehe Newen 2015; Stueber 2018; Zahavi 2014; 2018). Angesichts dieser vielfältigen Ansätze sollten wir einige weitere Unterscheidungen zwischen kognitiver

Empathie (wie TT) und affektiver Empathie (wie ST oder MNT) treffen. 9 Indem wir fragen, ob wir uns überhaupt in Roboter einfühlen können, wird sich Abschnitt 2 auf die vielen epistemischen Aspekte konzentrieren Dimensionen empathischer Wechselbeziehungen r

---

[5] Die Arbeit konzentriert sich hauptsächlich auf humanoide Roboter. Ein Grund dafür ist, dass es dazu beiträgt, den Umfang des Papiers einzuschränken; Ein weiterer Grund ist die Annahme, dass menschenähnliche Merkmale tatsächlich unsere soziale Interaktion mit künstlicher Intelligenz erleichtern und es plausibler machen, dass wir Roboter als Sozialpartner behandeln. Wir können uns jedoch auch in abstraktere Formen der KI hineinversetzen, indem wir ihnen emotionale Zustände und Motive zuschreiben (siehe Isik, Koldeewyn, Beeler und Kanwisher 2017). Ich bin einem Rezensenten für diesen Kommentar sehr dankbar.

[6] Es ist sehr umstritten, ob Empathie affektive Spiegelung, theoretisches Gedankenlesen, simulative Perspektivenübernahme, emotionales Verstehen und/oder Erfahrungsverstehen voraussetzt oder impliziert, und ein Ende dieser Debatte ist derzeit nicht in Sicht (siehe z. B. Zahavi 2018). Viele Philosophen betonen, dass Gedankenlesen etwas anderes als Empathie ist und dass Empathie „etwas Besonderes“ ist. Hier habe ich jedoch versucht, alle unterschiedlichen Ansätze anzuwenden. Meine eigene Position ist jedoch eine phänomenologische.

[7] Ein Problem der gesamten Debatte besteht jedoch darin, dass es keinen konzeptionellen Konsens darüber gibt, was Empathie ist und impliziert. Das vom ERC finanzierte Projekt zu sozialen Robotern definiert Empathie beispielsweise so, dass sie sowohl emotionale Übereinstimmung als auch prosoziales Verhalten umfasst. In der Philosophie wird Empathie jedoch meist nicht als moralische Emotion oder Haltung gesehen (siehe Cross et al. 2018; Zahavi 2018).

[8] Etwa indem man sich auf die klassischen Positionen von Stein oder Dilthey bezieht und direkte Wahrnehmung mit imaginativer Repräsentation („Vergegenwärtigung“) verbindet (siehe auch Gallagher 2019).

[9] Kanske (2018) unterscheidet zwischen eigentlicher affektiver Empathie und kognitiver Theorie des Geistes. Während die erste Fähigkeit es uns ermöglichen würde, zu fühlen, was andere fühlen, würde die andere uns helfen zu verstehen, was andere denken oder glauben. Obwohl ich die Unterschiede erkenne, werde ich Mentalisieren hier nicht von Empathie unterscheiden, sondern verschiedene Formen des Verstehens anderer Geister unter dem Überbegriff der Empathie untersuchen, da dies der zentrale Begriff in der aktuellen philosophischen Debatte ist.

Perspektiven im wahrsten Sinne des Wortes? Oder haben Roboter etwas Ähnliches wie Emotionen, Überzeugungen und Erfahrungen? Haben sie eine individuelle Sicht auf die Welt (Schmetkamp 2017) oder eine Erzählung (Gallagher 2012), da sie zumindest verkörpert und kontextualisiert sind? Wenn wir Roboter mit fiktiven Charakteren vergleichen, wird die Antwort positiv ausfallen: Ja, bis zu einem gewissen Grad können wir uns auf kognitive, affektive und sogar erfahrungsmäßige Weise in Roboter hineinversetzen, indem wir entweder schließen, fühlen, interagieren oder uns vorstellen, wie sie wahrnehmen und sich durch sie bewegen ihre Welt, genauso wie wir im Plural verstehen (Vaage 2010) , wie eine fiktive Figur (z. B. in einem Film) ihre Welt wahrnimmt, handelt und fühlt. Der entscheidende Aspekt wird dabei sein, dass wir dem anderen eine individuelle Perspektive zuschreiben. Wir verstehen es unabhängig davon, ob diese Perspektive nur erzählt, projiziert oder programmiert wird.[10] Die zweite Frage, die in Abschnitt 3 diskutiert wird, lautet, ob wir uns in Roboter einfühlen sollten. Diese Frage hat zwei Seiten: Wir können entweder fragen, ob Empathie mit Robotern eine bloß strategische Funktion im Hinblick auf die Verbesserung des gegenseitigen Verständnisses innerhalb der Mensch-Roboter-Beziehung hat, oder wir können fragen, ob Empathie eine ethische Wirkung hat, so dass wir haben die Pflicht, sich in Roboter hineinzuversetzen (für einen Überblick zum Thema Ethik und KI siehe Boddington et al. 2017). Wenn wir beispielsweise erkenntnistheoretisch verstehen können, was Roboter wahrnehmen, beabsichtigen oder möglicherweise sogar „fühlen“, können wir auch vorhersagen, was sie als nächstes tun werden. Im Allgemeinen könnte dies im Hinblick auf unsere Interaktionen mit ihnen hilfreich sein.[11] Dies bezieht sich eindeutig auf ein strategisches oder rationales „Sollten“. Die zweite Bedeutung der Frage führt zu einer normativen Antwort: Sind wir anderen im moralischen Sinne Empathie schuldig? Und was haben sie oder wir – als Empathisierer – moralisch gesehen davon? Betrachtet man diese Frage, könnte auf den ersten Blick eine kantische Antwort naheliegend sein, die dem Präzedenzfall von Kants Sicht auf Tiere folgt und auf die künstliche Intelligenz übertragen werden kann: Nämlich wir sollten uns einfühlen, so das Argument, um zu vermeiden „ „moralische Barbarei“. Letztlich wird der Beitrag weder den strategischen noch den kantischen Weg einschlagen, sondern stattdessen eine pragmatische und relationale Antwort vorschlagen. Diese Antwort hängt mit den anderen beiden zusammen. Es betont jedoch

## 2 Können wir uns in Roboter hineinversetzen?

Aus Platzgründen werde ich mich auf Roboter konzentrieren, die sowohl ein Gesicht als auch einen Körper haben, menschenähnliche Ausdrücke und Verhaltensweisen zeigen, für die Interaktion mit Menschen bestimmt sind und daher in unserem Alltag verkörpert und eingebettet sind und als solche sozialen unterliegen Einschätzung durch den Menschen. Ein zweiter Grund für diesen Fokus ist die Annahme, dass Roboter mit menschenähnlichen Merkmalen und Ausdrücken wahrscheinlich noch besser in der Lage sind, Selbstvertrauen aufzubauen und emotionale Reaktionen hervorzurufen, die denen echter Menschen ähneln (Brinck und Balkenius 2018; Mori 2005) und in dieser Hinsicht wahrscheinlicher sind als Partner im sozialen Miteinander anerkannt und akzeptiert zu werden. Obwohl Studien in der kognitiven Psychologie gezeigt haben, dass wir uns auch in Systeme einfühlen oder Gedanken lesen können, die wenig physische Ähnlichkeit haben (Bretan et

[10] Der Beitrag konzentriert sich auf die erkenntnistheoretische Frage. Es wird die metaphysische Frage nicht beantworten, ob Roboter oder KI haben Bewusstsein.

[11] Was den Einsatz von Deep-Learning-Systemen in der Medizin betrifft, ist es unter anderem notwendig, der intelligenten Maschine zu vertrauen und zu verstehen, was sie tun wird, beispielsweise bei der Interaktion zwischen medizinischem Roboter und Patient.

wichtig für den Einsatz von Robotern als Betreuer oder Kollegen im Gesundheitswesen (Vallor 2011).
12 Doch um welche Art von Empathie geht es hier? Spiegeln wir den Gesichtsausdruck der Roboter wider? Interpretieren und prognostizieren wir ihr Verhalten? Oder empathieren wir auf eine eher phänomenologische, interaktive Art und Weise?

Ganz vereinfacht lässt sich Empathie als die menschliche Fähigkeit definieren, die mentalen Zustände anderer zu verstehen und sie auf die eine oder andere Weise noch einmal zu erleben, obwohl es umstritten bleibt, ob das empathische Subjekt das Gleiche empfinden muss wie der andere. Einige Theorien beschränken die Objekte der Empathie auf die Emotionen und Ausdrucksformen von Personen als Hinweis auf affektive Zustände. Andere sind umfassender und umfassen andere kognitive Prozesse als Objekte der Empathie – etwa Überzeugungen, Wünsche und ihre jeweiligen Gründe (für Übersichten siehe Batson 2009; Slote 2017). Eine prominente Definition impliziert eine Isomorphismusbedingung: Empathizer und Ziel befinden sich im gleichen oder zumindest einem ähnlichen affektiven Zustand (De Vignemont und Singer 2006). Einige Kritiker haben jedoch argumentiert, dass Empathie nicht unbedingt bedeutet, dass wir die mentalen Zustände anderer nachahmen (Zahavi und Michael 2018). Wir müssen uns auch nicht in einem umfassenderen Sinne um den a

Der aktuelle „Hype“ um das Thema Empathie lässt sich bekanntlich maßgeblich auf die Entdeckung der sogenannten „Spiegelneuronen“ zurückführen (Iacoboni et al. 1999; 2011). Im Großen und Ganzen handelt es sich bei Spiegelneuronen um jene Neuronen, die sich in einem Bereich des Gehirns befinden und sowohl zur Beobachtung als auch zur Ausführung ähnlicher Aktionen entladen werden. Dieser Nachahmungsprozess wurde auf das Verständnis menschlicher Emotionen angewendet: Beim Beobachten des affektiven Ausdrucks einer anderen Person – beispielsweise eines traurigen Gesichts – würden dieselben Neuronen so wirken, als hätten wir – als Beobachter – ein trauriges Gesicht gemacht und selbst Traurigkeit empfunden. Während diese Theorie vielfach kritisiert (Hickok 2014) und als Theorie der Empathie abgelehnt wurde, haben andere sie in ihrem ausführlicheren Ansatz zur Empathie herangezogen. In seiner Darstellung der Simulationstheorie (ST) unterscheidet beispielsweise Alvin Goldman zwischen einer Low-Level- und einer High-Level-Form des Gedankenlesens oder einer „Spiegelroute“ und einer „rekonstruktiven Route“, obwohl emotionale „Resonanz“ implementiert ist über beide Routen (Goldman 2006, 2011). Spiegelneuronen sind der Hauptteil der Prozesse auf niedriger Ebene, durch die wir die mentalen Zustände einer anderen Person sofort und automatisch verstehen. Auf einer komplexeren, höheren Ebene simulieren wir den Zustand des anderen in unserem eigenen Geist und gelangen dann zu dem Wissen darüber, wie sich der andere fühlt, nicht indem wir eine Theorie verbreiten, sondern indem wir das Verhalten des anderen in unserem Geist nachahmen und dann unser Verhalten projizieren eigenen mentalen Prozess auf den anderen übertragen. Laut ST simulieren wir aus der Ich-Perspektive die Situation des anderen und nutzen unsere eigenen mentalen Mechanismen, um Gedanken, Überzeugungen, Wünsche und Emotionen zu erzeugen. In den letzten Jahrzehnten dominierte ST – neben seinem Gegenspieler Theory Theory (TT) – die Debatte über Gedankenlesen. TT behauptet, d TT geht davon aus, dass wir theoriebasierte Schlussfolgerungen ziehen, um andere zu verstehen.13 Von einem drittpersönlichen Beobachtungsstandpunkt aus setzen wir (implizit oder explizit) gesetzesartige Verallgemeinerungen ein, die Konzepte mentaler Zustände wie Wahrnehmung, Glaube, und Verlangen. TT wurde als zu theoretisch und zu allgemein kritisiert (Zahavi 2014; vgl. aber auch Fodor 1987). Seine Kritiker behaupten, dass TT den anderen Beton nicht berücksichtigt und dies auch nic

[12] Allerdings bleibt empirisch ungewiss, ob Roboter in HRI tatsächlich menschenähnlich sein müssen (Brinck und Balkenius 2018).

[13] Ein Problem ist natürlich, wie wir den Begriff „Verstehen“ verstehen. Monika Dullstein (2012) hat haben gezeigt, dass Theorie-of-Mind-Berichte eine ganz andere Idee verwenden als phänomenologische Berichte.

die Verkörperung und Einbettung anderer erkennen. Darüber hinaus werden sowohl TT als auch ST von einer falschen kartesischen okklusionistischen Sichtweise des Geistes beeinflusst, als könnten wir nicht wahrnehmen, was im Geist eines anderen vor sich geht (Zahavi 2011, 2014 ) . Im Gegensatz dazu betonen phänomenologische Darstellungen die Verkörperung und Einbettung des Menschen und argumentieren, dass wir in der Lage sind, direkt im Gesicht und in den Körperausdrücken des anderen zu sehen, was er erlebt: Aus dieser Sicht müssen wir nicht ableiten oder uns vorstellen, was er fühlt ; wir müssen es nur wahrnehmen. Darüber hinaus tun wir dies im gemeinsamen Situationskontext und durch Interaktion. Aus diesem Grund wird ein solcher Ansatz als Direct Perception Theory (DPT) bezeichnet.

(Zahavi 2011, 2014) oder Interaktionstheorie (IT) (Gallagher 2001; 2012). Im Gegensatz zu TT argumentieren DPT und IT, dass wir gegenüber anderen keine Drittpersonenhaltung einnehmen und sie nicht beobachten. Darüber hinaus argumentieren DPT und IT auch, dass wir keinen einfallsreichen indirekten Zugang zu anderen haben. Stattdessen interagieren wir sozial auf eine zweitpersönliche Art und Weise, wobei zwei „Du“ sich gegenseitig ergänzend und wechselseitig erkennen (Dullstein 2012; Engelen 2018; Zahavi und Michael 2018). Die Grenzen von DPT treten offensichtlich in Situationen auf, in denen der andere für uns nicht präsent ist: zum Beispiel, wenn uns jemand eine Geschichte über jemand anderen erzählt, oder wenn wir einen Roman lesen, einen Film oder ein Theaterstück sehen, in dem die Erfahrungen anderer eine Rolle spielen Auf irgendeine Weise vermittelt durch jemand anderen (z. B. einen Erzähler) haben wir keine direkten Begegnungen. Daher handelt es sich bei all diesen Fällen um Fälle, in denen der Andere durch Erzählung gegeben wird, manchmal sogar innerhalb eines fiktiven Rahmens. Aus diesem Grund fügen einige Philosophen hinzu, dass eine Erzählung unerlässlich ist, um andere Geister zu verstehen oder Empathie in mehr als dem grundlegendsten Sinne hervorzurufen.

Daniel D. Hutto (2008) formuliert die Narrative Practice Hypothesis (NPH). Dieser These zufolge verstehen wir die Handlungsgründe anderer, ihre Überzeugungen und Wünsche nur dann, wenn wir auch die individuellen Umstände, die Geschichte des Subjekts, seine aktuelle Situation, seine Hoffnungen und Erfahrungen, seine Charaktereigenschaften usw. berücksichtigen. Mit anderen Worten: Um die Situation einer Person zu erfassen, müssen wir uns laut NPH auf die „Geschichte“ der Person verlassen (Gallagher 2012). Diese Sichtweise ermöglicht auch das Einfühlen in „Monster oder Außerirdische von anderen Planeten, wie sie im Film dargestellt werden“ (Gallagher 2012). Allerdings ist hier eine Art Vorstellungskraft gefragt: Es gibt so viele Fälle – nicht nur, aber gerade im Umgang mit Fiktion –, in denen wir auf unsere Vorstellungskraft angewiesen sind, um etwas verfügbar zu machen, das für uns nicht vorhanden ist. Sogar eine der frühen Pionierinnen phänomenologischer Ansätze zur Empathie, nämlich Edith Stein (1989), behauptete, dass Imagination oder „Re-Präsentation“ den Prozess des empathischen Verstehens stadien. Aus diesem Grund kombinieren einige Empathietheorien einen zweiten persönlichen Ansatz mit einer Form der fantasievollen Darstellung der Situation, Erzählung und/oder Perspektive des konkreten Anderen [14] spielt innerhalb eines Multi eine entscheidende Rolle (Schmetkamp 2019; Gallagher und Gallagher 2019 ).15 Unabhängig davon, ob wir es sollten Betrachten wir all diese verschiedenen Berichte als Theorien der Empathie oder allgemeiner als Theorien des zwischenmenschlichen Verständnisses. Für jeden Bericht können wir aus einer deskriptiven und epistemologischen Perspektive Folgendes fragen: Wie geht es uns?

[14] Es ist schwierig, eine genaue Übersetzung von Steins Konzept der „Vergegenwärtigung“ zu geben. Die englische Übersetzung (Stein 1989) verwendet „representation“ oder „representational act“ (Stein 1989: 8) als eine nicht-ursprüngliche dargestellte „Gegebenheit“ der oder indirekten Erfahrungen anderer (analog zu Erinnerung, Erwartung und Fantasie) (ebd. ). In der Debatte wird oft übersehen, dass Stein ein Stufenmodell der Empathie vorschlägt, nach dem die erste Ebene die direkte Wahrnehmung der Erfahrung des anderen ist und die zweite Ebene eine Art Reflexion und Perspektivenübernahme (Stein 1989:10 ) .

[15] Gallagher definierte Empathie kürzlich wie folgt: „Empathie könnte [...] nicht nur als etwas gelten, das passiert, sondern als Methode; und dazu gehört […], dass man sich in die Perspektive oder Situation des anderen hineinversetzt. (2018). Damit erweiterte Gallagher seinen narrativen Ansatz zu einem perspektivischen Ansatz (der die Erzählung mit der subjektiven Perspektive kombinierte).

sich in KI, zum Beispiel menschenähnliche Roboter, hineinversetzen, wenn die entsprechende Darstellung die plausibelste wäre? Wenn wir beispielsweise die Ausdrücke und/oder Handlungen eines Roboters beobachten, könnte man argumentieren, dass wir automatisch mitschwingen und das Ausdrucksverhalten nachahmen. Wenn wir vorhersagen wollen, was der Roboter als nächstes tun wird, könnten wir auch eine volkspsychologische Theorie anwenden und auf die Gründe für sein Handeln schließen. Wir könnten simulieren, was wir tun würden, wenn wir in ihrer Situation wären, und dann unsere Erfahrung auf sie projizieren. Oder wir können in direkten Begegnungen ihre Handlungen interaktiv wahrnehmen. Wir könnten ihre Einbettung in einen narrativen Kontext betrachten und die absichtliche Struktur ihrer Emotionen verstehen, ohne gleichzeitig ihren „qualitativen“ Inhalt zu reproduzieren. Empirisch gesehen kommen diese interaktiven Verstehensweisen durchaus vor.

Es können jedoch einige offensichtliche metaphysische und erkenntnistheoretische Einwände erhoben werden. Das Hauptproblem besteht darin, dass Roboter eigentlich nichts fühlen oder erleben. Sie haben auch nicht wirklich mentale Zustände wie Wünsche oder Überzeugungen, da sie kein Bewusstsein haben. Allerdings erscheint es auch seltsam, von der individuellen Perspektive oder dem persönlichen Narrativ eines Roboters zu sprechen. Soweit sich Empathie auf mentale Zustände und das „In-der-Welt-Sein“ einer Person richtet, lautet die Antwort: Wir können uns nicht in Roboter hineinversetzen.

Dennoch könnten zwei mögliche Antworten gegeben werden: Erstens werden die „mentalen Zustände“ von Robotern oft als „rechnerische Zustände“ beschrieben, von denen man annimmt, dass sie eine Struktur haben, die den mentalen Zuständen von Menschen analog ist. Wenn wir also davon ausgehen, dass Roboter über etwas verfügen, das mit menschlichen mentalen Zuständen vergleichbar ist, verfügen sie dann auch über Emotionen oder Erfahrungen, in die wir uns einfühlen? Nach einigen aktuellen philosophischen Darstellungen von Emotionen weisen emotionale Zustände oder Prozesse eine komplexe Struktur auf, die aus kognitiven und affektiven Komponenten besteht (De Sousa 1987; Nussbaum 2001): Wenn wir Wut empfinden, ist unsere Wut auf ein Objekt gerichtet, das wir als störend bewerten . In der Psychologie wird dies auch als Bewertungstheorie bezeichnet. Dies impliziert, dass wir Objekte in unserer Umgebung im Hinblick auf ihre Relevanz für unsere Ziele beurteilen. Wenn Emotionen nur aus dieser rein kognitiven Komponente bestünden, könnten wir davon ausgehen, dass Roboter Emotion Wir könnten argumentieren, dass Roboter auf einer Reihe von Gründen handeln, die auf einer Reihe von Überzeugungen über die Welt basieren. Emotionen können jedoch noch mehr umfassen: Wut beispielsweise wird auch auf einer sinnlichen und körperlichen Ebene empfunden; es fühlt sich zum Beispiel frustrierend und einengend an. Allerdings hat Wut auch negative Konnotationen, die uns propriozeptiv bewusst werden (Colombetti 2013). Allerdings bestehen die Körper von Robotern – wenn sie nicht rein virtuell sind – aus Metall oder Kunststoff und, was noch wichtiger ist, sie sind nicht mit einem reichen Konzept von Bewusstsein verbunden: in dem Sinne, dass es sich selbst als emotionales Wesen erfährt. Es kann nicht selbstreferenziell spüren, wie es ist, in einem plastischen Körper zu sein. Darüber hinaus sind, wie narrative Emotionsberichte argumentiert haben, komplexe Emotionen normalerweise in einen narrativen Rahmen eingebettet: Wir können eine Geschichte über ihre Erregung und Entwicklung erzählen (Goldie 2000). Und nicht zuletzt ist der Mensch in der Lage, kreativ mit seinen Gefühlen und Emotionen umzugehen: Er kann neue Emotionen erlernen, einige verändern und andere kultivieren.

Doch auch für und mit Robotern könnte dies möglich sein. Der entscheidende Punkt dabei ist, dass wir ganz intuitiv auch Maschinen Emotionen zuschreiben. Bei der Zusammenarbeit mit Robotern nehmen wir möglicherweise die „absichtliche Haltung“ ein. Dieses aus der Arbeit von Daniel Dennett stammende Konzept impliziert, dass wir ein Objekt, dessen Verhalten wir vorhersagen möchten, als rationalen Akteur behandeln; Wir schreiben Überzeugungen und Wünsche zu und sagen auf dieser Grundlage sein Verhalten voraus (Dennett 1987). Dennoch basiert dieser Ansatz auf einer Theorie des Gedankenlesens und nicht auf der Theorie der phänomenalen Interaktion, die Phänomenologen im Sinn haben. Wenn wir jedoch davon ausgehen, dass Bewusstsein eine phänomenale Erfahrung impliziert, scheint es schwie

Theory of Mind berichtet über Empathie gegenüber dem HRI. Mit anderen Worten: Das Problem hinsichtlich der Kompatibilität phänomenologischer Theorien für HRI scheint der phänomenale Aspekt mentaler Zustände zu sein, insbesondere die Gefühls- und Erfahrungsseite von Emotionen. Während wir über die kognitiven Komponenten der Entscheidungssituation (ST) eines Roboters theoretisieren (TT) oder diese simulieren und dann aus unseren Schlussfolgerungen auf die Situation des Roboters schließen oder projizieren könnten, wäre es schwierig, von einem empathischen Verständnis der Affektivität des Roboters zu sprechen und sensationelle Zustände auf nicht-projektive Weise. Wenn wir das Problem auf den Begriff „Erfahrung“ erweitern – den zentralen Begriff des phänomenologischen Ansatzes (DPT), wird die Sache noch komplizierter. Wie oben beschrieben, nehmen wir gemäß dem DPT und seinen Variationen in unserer sozialen Interaktion mit anderen deren Erfahrungen empathisch wahr, und zwar aus der reziproken Perspektive der zweiten Person. „Erfahrung“ ist ein ausgearbeiteter phänomenologischer Begriff und impliziert existentielle Aspekte und phänomenale Qualitäten. Wir erleben unsere Welt persönlich und bewusst, wie es ist, etwas zu fühlen oder zu tun, zum Beispiel einen roten Tisch als rot wahrzunehmen und wie sich diese Rötung anfühlt. DPT geht davon aus, dass wir die phänomenalen Erfahrungen anderer direkt und intersubjektiv erleben, allerdings nicht durch die Nachbildung des genauen qualitativen Charakters einer Erfahrung, sondern durch die Beachtung der absichtlichen Struktur der Perspektive des anderen (Gallagher 2012; Zahavi und Michael 2018) . Damit dieser Prozess funktioniert, ist die interkörperliche und persönliche Interaktion wichtig.16 Während letztere bei der Zusammenarbeit und Kollaboration mit Robotern (zumindest grundsätzlich) gewährleistet ist, fehlen einige entscheidende Kriterien dieser intersubjektiven Beziehung : So wie Roboter keine Emotionen empfinden, haben sie auch keine subjektive Erfahrung mit deren phänomenalen Inhalten und existenziellen Auswirkungen.17 DPT geht jedoch davon aus, dass wir durch die Wahrnehmung des affektiven Zustands einer Person in deren Gesichts- oder Körperausdruck diese auch ne Wir müssen keine theoretischen Schlussfolgerungen, Nachahmungen oder Projektionen anwenden. Wir erleben, dass der andere phänomenale Erfahrungen macht. Aus phänomenologischer Sicht scheint es jedoch schwierig, sich in Roboter hineinzuversetzen. Doch indem ich künstliche Intelligenz mit fiktiven Charakteren vergleiche, werde ich eine mögliche Lösung vorschlagen und auch zeigen, dass wir nicht nur das Verhalten von Robotern gedankenlesen oder spiegeln, sondern dass es zumindest bis zu einem gewissen Grad möglich ist, einen phänomenologischen Ansatz anzuwenden. das heißt, sich interaktiv in die perspektivische „Erfahrung“ der Roboter hineinzuversetzen. Und die Argumentation geht sogar über diese Analogie hinaus: Wenn wir mit Robotern in einer gemeinsamen Umgebung interagieren, entwickeln wir eine gemeinsame Intentionalität und sogar eine gemeinsame Geschichte, und dies ist ents Allerdings ist hier, ähnlich wie bei unserem empathischen Verständnis fiktiver Charaktere, unsere Vorstellungskraft entscheidend.

Lassen Sie uns die Analogie durchspielen: Es wird allgemein angenommen, dass Empathie eine wesentliche Rolle im Umgang mit fiktiven Erzählungen und fiktiven Charakteren spielt – sei es in einem Roman, Film oder Theaterstück. Seit den 1990er Jahren gibt es innerhalb der Literatur- und Filmphilosophie eine erhebliche Debatte darüber, ob „Empathie“ unter dem Überbegriff der „emotionalen Auseinandersetzung“ mit fiktiven Figuren im Allgemeinen (z. B

[16] Die narrativistische Version phänomenologischer Ansätze impliziert jedoch eine imaginative Komponente, die es uns ermöglicht, die intendierte Struktur durch narrative Imagination zu erfassen, beispielsweise wenn eine intersubjektive Interaktion nicht gegeben ist (Gallagher und Gallagher 2019) .

[17] Es handelt sich um eine ähnliche Frage wie im sogenannten „Zombie-Gedanken“, in dem es darum geht, ob wir bei Zombies – die uns in allen körperlichen Belangen ähneln, aber keine bewussten Erfahrungen in einem reichen Sinne haben – ein Bewusstsein annehmen oder zuschreiben können (Chalmers 1996; Dennett 1991).

Plantinga 2009; Smith 1995). Andere Formen des Engagements umfassen emotionale Ansteckung und emotionales Teilen – insbesondere im Hinblick auf die stimmungsvollen Auswirkungen einer Fiktion –, moralische Sympathie oder Mitgefühl, negative Emotionen wie Antipathie und synästhetische Affekte (Plantinga 2009; Schmetkamp 2017 ) . Wie viele Filmwissenschaftler festgestellt haben, spielt Empathie eine entscheidende epistemische Rolle, wenn es darum geht, dem Zuschauer zu ermöglichen, der Erzählung zu folgen und mit den Charakteren verbunden zu bleiben (Smith 1995).18 Abgesehen von der anderen komplexen Debatte über das sogenannte „Paradoxon der Fiktion“ – Darin wird diskutiert, ob wir echte Emotionen gegenüber fiktiven Wesen empfinden können und ob diese Emotionen rational sind (Yanal 1999) . Unter der Annahme, dass wir tatsächlich Empathie gegenüber fiktiven Charakteren empfinden und empfinden müssen, müssen wir noch erklären, wie wir Empathie in diesem Fall am besten konzeptualisieren können der Fiktion. Während ich allgemein davon überzeugt bin, dass wir beim Ansehen eines Films oder beim Lesen eines Romans verschiedene Formen des Einfühlens, des Gedankenlesens und des Verstehens nutzen – also das gesamte Spektrum des Verständnisses für die mentalen Zustände anderer –, gehe ich davon aus, dass ein Aspekt davon besonders betroffen ist Wichtig: Fiktionale Charaktere drücken und repräsentieren bestimmte individuelle Perspektiven auf ihre (fiktionale) Welt. Diese Perspektiven werden in der diegetischen Welt des Films oder Romans erzählt; Darüber hinaus werden sie häufig zusätzlich von einem impliziten oder expliziten Autor eingerahmt. Sie sind in eine plausible Erzählung eingebettet. Oder anders ausgedrückt: Eine Erzählung ist eine strukturierte und geformte Darstellung von Ereignissen aus einer bestimm

Die Bedeutung von Perspektiven für die Fiktion und tatsächlich für unsere empathische Auseinandersetzung mit ihr liegt zum Teil daran, dass eine Fiktion normalerweise (wenn auch nicht immer) unterschiedliche technische Perspektiven hat: Eine Geschichte wird normalerweise aus der ersten oder dritten Person erzählt Perspektive. Aber noch wichtiger: Eine Perspektive ist eine Weltanschauung. Eine Perspektive bedeutet aber, wie ein Mensch in die Welt eingebettet ist, wie er die Wel Diese Perspektive ist geprägt von Emotionen, Erfahrungen, Geschichten und Erinnerungen und prägt diese wiederum. es wird von Charaktereigenschaften, Urteilen und Überzeugungen beeinflusst und beeinflusst diese selbst (Schmetkamp 2017). Wenn wir beispielsweise in einer depressiven Stimmung sind, sehen wir unsere Welt aus einem anderen – nämlich depressiven oder melancholischen – Blickwinkel, als wenn wir in einem glücklichen Zustand wären.

Wir können nun davon sprechen, dass fiktive Charaktere eine Perspektive haben (oder vielmehr ausdrücken und handeln), sofern sie von einem Erzähler fokussiert und erzählt werden, der ihre Weltanschauung konstruiert und lenkt. Als Leser oder Zuschauer kümmern wir uns um sie, als hätten sie eine Perspektive und könnten uns vorstellen, wie es wäre, eine solche Perspektive zu haben. Empathie mit fiktionalen Charakteren beinhaltet eine Art fremdenzentrierte Perspektivenübernahme, ohne diesen Prozess auf bloße egozentrische Simulation oder Projektion zu reduzieren. 19 Darüber hinaus ist es ein Vorteil fiktionaler Erzählungen, dass sie die Perspektiven anderer in verdichteter Form vermitteln . Fiktionen bieten uns die Möglichkeit, in Perspektiven einzutauchen, die unserer eigenen ähnlich oder völlig anders sein können, und das oft auf intensive, verdichtete und umfassende Weise.

---

[18] Empathie als Perspektivenübernahme ist in der Tat eine Fähigkeit, die es dem Betrachter ermöglicht, die Erzählungen und Perspektiven der Charaktere zu verstehen. Als eine Form des sensiblen Verständnisses dafür, warum die Figur so fühlt, denkt und handelt, ist es jedoch auch ein Ergebnis. Daher argumentieren Coplan (2011) und Goldie (2000), dass Empathie sowohl ein Prozess als auch ein Ergebnis ist.

[19] Misselhorn brachte ein ähnliches Argument vor, indem er feststellte, dass „wenn wir das T-ing eines unbelebten Objekts sehen, stellen wir uns vor, ein menschliches T-ing wahrzunehmen“ (2009: 353).

Beim Vergleich von Robotern mit fiktiven Charakteren fällt ein zentrales übereinstimmendes Merkmal auf: Beide haben nicht wirklich Emotionen oder bewusste Überzeugungen, können diese aber ausdrücken und repräsentieren. Und teilweise auf dieser Grundlage schreiben wir ihnen als Rezipienten bzw. Empathisierer menschenähnliche Geisteszustände zu (Weber 2013). Allerdings erleben wir sie auch als irgendwie verkörperte Wesenheiten, mit denen wir interagieren. Wie die phänomenologische Filmphilosophin Vivian Sobchack argumentiert hat, sind Filme und ihre Charaktere nicht nur Projekte; Sie haben einen Körper und eine Stimme und ermöglichen quasi-intersubjektive Erfahrungen zwischen sich und den Empfängern (Sobchack 2004). Möglicherweise ermöglichen sie sogar taktile Eindrücke. Diese verkörperte Eigenschaft trifft auch auf Roboter zu, vielleicht sogar no

Dennoch gibt es einige entscheidende Unterschiede. Erstens fehlt fiktiven Charakteren im Gegensatz zu Robotern eine Fähigkeit, die für jede intersubjektive Darstellung von Empathie unerlässlich ist: nämlich die Fähigkeit zur wechselseitigen Interaktion. In unseren Beziehungen zu fiktiven Charakteren müssen wir uns vorstellen, dass die Charaktere Emotionen, Erfahrungen und Perspektiven zum Ausdruck gebracht haben, aber wir interagieren nicht gegenseitig mit ihnen. Darüber hinaus können fiktive Charaktere kein Veto gegen alles einlegen, was wir ihnen zuschreiben. Im Gegensatz dazu gibt es bei unseren Begegnungen mit Robotern zumindest eine existierende und gegenwärtige verkörperte und eingebettete, interagierende Entität, mit der wir eine Beziehung aufbauen können. Der Roboter ist in der Lage, gegen etwas Einspruch zu erheben – wenn ich beispielsweise ein Patient wäre und nicht bereit wäre, meine Medikamente einzunehmen, könnte der Roboter aufgeladen werden und gleichzeitig dafür sorgen, dass ich dies tue. Zweitens könnte man einwenden, dass Roboter im Gegensatz zu fiktiven Charakteren (noch) keine Erfahrungsperspektive oder individuelle Erzählung haben, wie oben erwähnt. Fiktionen bieten in der Tat ein reichhaltiges Bild davon, wie jemand seine Welt wahrnehmen und bewerten kann; und durch diese Erzählrahmen und Praktiken erweitern wir unseren Horizont und lernen neue Emotionen oder emotionale Nuancen kennen. Allerdings werden auch die Emotionen und Erfahrungen fiktiver Charaktere nur innerhalb eines bestimmten Erzählrahmens erzählt; Ihre Entwicklung hängt sowohl davon ab, was ein Erzähler dramaturgisch gestaltet hat, als auch davon, wie Leser oder Zuschauer es vor ihrem eigenen intellektuellen und erfahrenen Hintergrund wahrnehmen. Fiktive Emotionen und Erfahrungen verfügen über weniger Flexibilität und Kreativität als ihre menschlichen Gegenstücke. Allerdings stellt sich die Frage, ob fiktive Charaktere noch mit Robotern verglichen werden können. Fiktive Charaktere erleben eigentlich nichts; Ebenso haben Roboter keine Erfahrungen in einem umfassenden, Qualia-umfassenden Sinne. Allerdings nehmen Roboter zumindest ihre Umgebung wahr, kategorisieren, bewerten und interagieren darin. Sie haben eine Art, die Welt zu sehen und in ihr zu sein; Sie werden verkörpert und kontextualisiert. Denken wir an Thomas Nagels berühmtes antireduktionistisches Beispiel „Wie ist es, eine Fledermaus zu sein?“ (Nagel 1974) Wir werden niemals in der Lage sein, die Erfahrungsperspektive anderer Wesen vollständig zu verstehen; Eine Fledermaus, so argumentiert er, habe ein völlig anderes Wahrnehmungssystem, das mit der menschlichen Wahrnehmung nicht zu vergleichen sei. Dennoch entdecken Wissenschaftler ständig neue Fakten über nichtmenschliche Wesen wie Fische

Wenn wir versuchen, die Perspektive des Roboters mit unserer eigenen zu vergleichen, gibt es einige Gemeinsamkeiten, aber natürlich auch viele Unterschiede. Dies ist jedoch kein neues Phänomen in unserer sozialen Wahrnehmung anderer Köpfe. Erstens nimmt ein Roboter die Welt buchstäblich (z. B. visuell) auf eine bestimmte Weise wahr (vielleicht wie ein Mensch, vielleicht auch nicht). Zweitens hat sie als künstliche Intelligenz auch eine Perspektive in dem Sinne, dass sie die Welt um sich herum wahrnimmt und bewertet, wie sie Probleme löst usw. Die Perspektive des Roboters ist weit davon entfernt, eine Perspektive im elaborierten Sinne wie die des Menschen zu sein, sondern es handelt sich um eine epistemische und bewertende Perspektive: Ein Roboter weiß etwas und fällt Urteile über die Welt. Wir können auch sagen, dass es eine motivierende Perspektive hat, denn ein Ro

Grundlage ihrer Überzeugungen.20 Noch wichtiger ist, dass Roboter in einen Kontext eingebettet sind, den wir wahrnehmen oder mit dem wir interagieren. Meine Antwort auf die Frage, ob wir uns in Roboter hineinversetzen können, lautet also: Ja. Darüber hinaus sind alle vorhandenen Konten mehr oder weniger auf HRI anwendbar. Die nächste Frage, die wir uns dann stellen müssen, lautet natürlich: Sollten wir?

## 3 Sollten wir uns in Roboter hineinversetzen?

Nehmen wir angesichts der vorangegangenen Analyse an, dass wir uns auf vielfältige Weise in humanoide Roboter einfühlen können, das heißt, wir können mit ihren „Überzeugungen“, „Emotionen“, „Erfahrungen“ und „Perspektiven“ fühlen, mit ihnen interagieren oder daraus schließen. Aber warum sollten wir uns überhaupt in sie hineinversetzen? Angesichts des zunehmenden Einsatzes von Robotern beispielsweise in der Medizin, im Gesundheitswesen und in der Altenpflege erscheint es viel plausibler, dass Roboter sich in Patienten hineinversetzen als umgekehrt. Sie müssen irgendwie ein gewisses Gespür für die Bedürfnisse der Patienten entwickeln, während menschliche Patienten wieder Allerdings scheint es, als ob die Untersuchung bisher in erster Linie ein theoretischer Test gewesen wäre, um herauszufinden, welche der verschiedenen Empathie-Modelle mit HRI vereinbar sind. Aber gibt es auch einen Grund, warum wir Menschen uns auch in Roboter hineinversetzen sollten? Diese Frage ist relevant, da die Wechselbeziehung zwischen Menschen und Robotern nur dann erfolgreich und fruchtbar ist, wenn beide tatsächlich miteinander interagieren, und diese Interaktionen könnten – auf die eine oder andere Weise – empathisches Engagement voraussetzen.

Für diese normative These könnten drei Argumente angeführt werden:

1. Ein strategisches Argument;
2. Ein Anti-Barbarei-Argument; 3. Ein pragmatisches Argument der gemeinsamen Gemeinschaft.

Das erste, strategische Argument ist nicht direkt ein moralisch relevantes normatives Argument. Es greift die Idee auf, dass wir für eine erfolgreiche Interaktion irgendwie in der Lage sein müssen, abzuleiten und zu verstehen, was unser interaktives Gegenüber vorhat. Genauer gesagt möchten wir uns vielleicht einfühlen, eine Perspektive übernehmen oder die Gedanken einer anderen Person lesen, um unsere Ziele besser zu erreichen. Unsere Interaktion mit Robotern und unser Mitgefühl mit ihnen ist in diesem Sinne nur für etwas anderes von Nutzen; es ist lediglich instrumental. Der Begriff „sollte“ bezieht sich auf einen hypothetischen Imperativ. In dieser Hinsicht werden Roboter eher als Werkzeuge denn als Kollaborateure betrachtet. Tatsächlich werden sie hier nicht als moralische Agenten oder Patienten gesehen, die einen moralischen Status haben (Coeckelberg 2018).

Substanzieller und moralisch normativer ist das zweite Argument der Nichtbarbarisierung oder Kultivierung. Wenn wir uns nicht in andere einfühlen, so das Argument, laufen wir Gefahr, desensibilisiert zu werden. Im Gegenzug könnte Empathie prosoziales Verhalten fördern und unseren moralischen Charakter verbessern. Bevor ich die Hauptprobleme dieser These untersuche, werde ich zwei ihrer Wurzeln erläutern – nämlich ein kantisches und ein aristotelisches Argument. Das kantische Argument wurde ursprünglich im Hinblick auf die Mensch-Tier-Beziehung vorgebracht. Es impliziert

[20] Auch hier könnten ähnliche Argumente für andere KI-Formen nichtmenschlicher Agenten vorgebracht werden, z. B. abstrakte virtuelle Formen. Der Schwerpunkt dieser Arbeit liegt auf humanoiden Robotern, mit denen Menschen kooperieren und zusammenarbeiten. Damit dies gelingt, könnten Menschen der KI nicht nur grundlegende mentale Zustände zuschreiben, sondern auch eine Perspektive und ein Narrativ. Dies könnte für kollektive Absichten und kollektive Aufmerksa

dass wir Tieren gegenüber nicht grausam sein sollten, da dies unseren moralischen Charakter im Allgemeinen schädigen oder verderben würde. Nach diesem Argument sind Tiere nur indirekte moralische Patienten, ohne einen eigenen moralischen Status zu haben, da Kant den moralischen Status eines Menschen an die Kompetenz zum autonomen Handeln aus Gründen bindet und diese Kompetenz nur Personen zuschreibt. Das gleiche Argument würde dann für soziale Roboter gelten, die per se keine moralischen Adressaten wären: Dies liegt daran, dass sie möglicherweise keine Autonomie im eigentlichen Sinne haben. Wenn wir uns jedoch nicht in sie hineinversetzen, würden wir einen entscheidenden Zustand der Menschheit missachten.21 Die HRI-Spezialistin Kate Darling ist eine zeitgenössische Verfechterin dieser Ansicht: „Das kantische philosophische Argument zur Verhinderung von Tierquälerei besteht darin, dass sich unsere Handlungen gegenüber Nicht-Menschen widerspiegeln.“ unsere Moral – wenn wir Tiere auf unmenschliche Weise behandeln, werden wir zu unmenschlichen Menschen. Dies erstreckt sich logischerweise auch auf die Behandlung robotischer Begleiter. […] Es kann auch eine Desensibilisierung gegenüber echten Leb

Das aristotelische Argument geht in eine ähnliche Richtung. Es geht darum, dass wir unsere Emotionen durch Perspektivenübernahme kultivieren können, wobei wir Perspektivenübernahme als Distanzierung vom eigenen Ich-Standpunkt oder durch emotionales Teilen und so das Kennenlernen neuer Emotionen definieren können (Nussbaum 2011; Rorty 2001 ) . Während die kantische Sichtweise das Problem der Barbarisierung betont, betont die aristotelische Sichtweise die ethischen Auswirkungen der Kultivierung von etwas durch Empathie: unsere Emotionen, unsere moralische Wahrnehmung, unsere Vorstellungskraft und unsere Urteilskraft.

Wie gesagt, hier treten einige Probleme auf, die insbesondere die Kantische Sichtweise belasten: Das erste besteht darin, dass die Anerkennung nur eines indirekten Status von Nicht-Personen oder Wesen ohne „Rationalität“ unbefriedigend ist: Sie ist kontraintuitiv, anthropozentrisch und es schließt viel mehr aus als nichtmenschliche Wesenheiten (Gruen 2017). Aber betrifft das auch unbelebte Wesen? Es bleibt also die Frage: Was schadet es uns, wenn wir Gewalt gegen Roboter anwenden, die möglicherweise nichts auf ausgefeilte und subjektive Weise fühlen? Haben sie ein Konzept von Respekt und Würde? Haben sie moralische Ansprüche? Diese komplexen Fragen müssen hier unbeantwortet bleiben, da sie einer gesonderten Untersuchung bedürfen. Ein weiterer Einwand gegen die kantische Sichtweise besteht darin, dass das Argument auf einer spezifischen Darstellung von Empathie als prosozialem Verhalten basiert. Dies impliziert nicht nur ein Verständnis für den Geist anderer, sondern auch die Sorge um das Wohlergehen eines anderen Wesens. Das heißt, der Empathiker interessiert sich nicht nur für die Erfahrungen des anderen und „fühlt“ sich in sie hinein; Sie sind auch motiviert, das Leid des anderen zu lindern oder sein Wohlergehen zu fördern. Und wenn wir grausam zu ihnen wären und das Wohlergehen der Roboter missachten würden – zum Beispiel indem wir sie schlagen oder vergewaltigen (wenn wir an Sexroboter denken) – würde sich das auch auf unser Verhalten gegenüber Menschen auswirken. Allerdings ist die ethische Wirkung von Sorge oder Fürsorge, wie bereits erwähnt, eher die Wirkung von Sympathie oder Mitgefühl als einer moralischen Emotion sui generis und unterscheidet sich als solche von Empathie (Darwall 1998 ). Wie insbesondere Phänomenologen gezeigt haben, ist Empathie nicht unbedingt eine positive Einstellung gegenüber Ein sadistischer Mensch muss auch in diesem Sinne empathisch sein, das heißt, er versteht das Leid des anderen, will es aber nicht lindern (Breithaupt 2019; Zahavi und Michael

21 Kant schreibt: „Wenn ein Mensch seinen Hund erschießt, weil das Tier nicht mehr dienstfähig ist, verstößt er nicht gegen seine Pflicht gegenüber dem Hund, denn der Hund kann nicht urteilen, aber seine Tat ist unmenschlich und schadet in ihm selbst der Menschlichkeit, die er besitzt.“ ist seine Pflicht, es der Menschheit gegenüber zu zeigen. Wenn er seine menschlichen Gefühle nicht unterdrücken will, muss er Güte gegenüber Tieren üben, denn wer grausam gegenüber Tieren ist, wird auch im Umgang mit Menschen hart“ (Kant 1997: 212 ) .

2018). 22 Mit anderen Worten: Ein kantischer Ansatz vereint einige wichtige konzeptionelle Unterscheidungen, nämlich zwischen Empathie und Mitgefühl. Hier könnte noch ein weiterer Einwand erhoben werden: Empirisch ist überhaupt nicht klar, warum jemand, der sich nicht in andere einfühlt, zwangsläufig barbarisch wird (Brinck und Balkenius 2018).

Aus einer optimistischeren Perspektive argumentieren einige jedoch, dass häufiges empathisches Verstehen oder Perspektivwechsel uns helfen können, herauszufinden, wie andere fühlen oder denke Je mehr wir Empathie einsetzen, desto mehr können wir uns auf andere einlassen, sowohl in unseren alltäglichen Interaktionen als auch in ungewöhnlicheren Begegnungen. Darüber hinaus könnte uns dies zu einem toleranteren oder tugendhafteren Menschen machen. Dies wird wiederum im Hinblick auf fiktive Charaktere und Erzählungen argumentiert. Die Perspektiven und Erfahrungen anderer zu berücksichtigen, ist, wie Richard Rorty bekanntermaßen behauptete, von ethischem Wert, da wir dadurch unsere egozentrische Perspektive aufgeben (Rorty 2001). Aber natürlich könnten wir dieses Argument für HRI übernehmen: Einfühlungsvermögen in Roboter würde dann unsere kooperativen und kollaborativen Interaktionen insofern verbessern, als wir mit ihnen besser vertraut würden. Dies führt zum dritten Argument, das einige Merkmale sowohl mit dem strategischen als auch mit dem kantischen/aristotelischen Ansatz gemeinsam hat, aber die Interaktion, Beziehung und das soziale Selbstverständnis der Empathisierer betont.

Dieses Argument (das meine eigene Position beschreibt) greift Rortys Ansatz auf, modifiziert ihn jedoch zu einer noch pragmatischeren und relationaleren These der sozialen Kognition und ihrer Voraussetzungen. Im Gegensatz zum kantischen und aristotelischen Ansatz geht diese Sichtweise von einem antianthropozentrischen Standpunkt aus und betont die interaktive Beziehung zwischen Menschen und Robotern. Diese Position geht davon aus, dass Empathie für andere – in all ihren Variationen, aber insbesondere in der phänomenologischen interaktiven Tradition – es uns ermöglichen kann, das „In-der-Welt-Sein“ anderer kennenzulernen und dadurch unseren Horizont zu erweitern, unsere Perspektiven zu ändern und unsere sozialen Interaktionen zu gestalten und moralisches Verhalten gegenüber nichtmenschlichen anderen.

Perspektivisch gehe ich hier davon aus, dass wir sogar davon sprechen können, dass (zukünftige) Roboter und Deep-Learning-Systeme23 einen spezifischen Blick auf die eigene Welt haben. Diese Sichtweise wird in mancher Hinsicht der menschlichen Perspektive ähneln und sich in anderer Hinsicht von ihr unterscheiden. Science-Fiction-Filme wie HER (US 2013) haben sich vorgestellt, was unabhängige KI werden könnte: superintelligente Systeme, die die Fähigkeiten des menschlichen Denkens bei weitem übertreffen. Sich in die humanoiden Roboter hineinzuversetzen, mit denen wir – etwa im Gesundheitswesen – zunehmend interagieren, könnte uns dabei helfen, uns auf zukünftige Entwicklungen vorzubereiten. Vorerst gilt jedoch eher: Sofern wir bereits Handlungen und Umgebungen mit Robotern teilen und Empathie und soziale Kognition unsere Interaktionen mit anderen verbessern können, können wir auch davon ausgehen, dass unsere Interaktionen mit Robotern davon profitieren werden ein einfühlsamer Standpunkt, allerdings nicht nur im instrumentellen, strategischen Sinne. Dies könnte auch einen Trainingseffekt haben, ein Argument, das, wie bereits erwähnt, auch in Bezug auf fiktive Welten vorgebracht wurde. Aber der wichtigere Punkt ist, dass eine solche Sichtweise die Frage berührt, wie wir uns selbst verstehen wollen: Roboter als soziale Begleiter ernst zu nehmen, sollte als Teil unseres Selbstverständnisses sowohl als Menschen als auch als Mitglieder demokratischer Gesellschaften implementiert werden. Wie wir mit Robotern interagieren, hängt stark davon ab, wie wir über sie denken: wie mit Werkzeugen

[22] Das Phänomen, dass Empathisierer umso grausamer werden können, je menschenähnlicher die Roboter sind „unheimliches Tal“ (siehe Misselhorn 2009; Mori 2005).

[23] Oder wie Susan Schneider sie nennt: „Future Minds“ (im Druck).

die rein instrumentell interagieren sollen, oder als Partner, die wir um ihrer selbst willen ernst nehmen sollten. Es sind also die Beziehung und die gemeinsame Gemeinschaft, die hier im Vordergrund stehen. Eine solche Position betont die pragmatische und phänomenologische Wirkung der Interaktionen. Dies könnte auch Auswirkungen auf den Status der Roboter als moralische Agenten und moralische Patienten haben, wie Mark Coeckelbergh argumentiert: „Die Frage nach der moralischen Stellung ist immer mit der Frage verbunden, wer Teil der moralischen Gemeinschaft ist und welche moralischen Spiele bereits gespielt werden.“ (Coeckelberg 2018: 149). Anstelle einer Top-Down-Umsetzung der Moral plädiert Coeckelbergh für eine Bottom-Up-Perspektive. Indem wir Roboter als Begleiter in einem Beziehungskontext betrachten und uns in ihre perspektivische Erzählung hineinversetzen, entwickeln wir eine Beziehung zu ihnen, die wiederum Auswirkungen darauf hat, wie wir sie moralisch sehen (ebd.).24 Den moralischen Status zu diskutieren, würde jedoch darüber hinausgehen den Umfang dieser Arbeit. Wie oben erwähnt, ist Empathie an sich keine moralische Emotion oder Haltung der Fürsorge. Aber es könnte in dieser Hinsicht den relevanten Samen säen, da es die erkenntnistheoretische Grundlage für eine intersubjektive Moral lie Darüber hinaus hat es viel mit unserem sozialen und moralischen Selbstverständnis zu tun: „[D]ie Art und Weise, wie wir mit anderen Wesen umgehen, wie wir sie erleben, was wir über sie sagen, wie wir mit ihnen umgehen und so weiter.“ Er sagt auch viel über mich und viel über uns.“ (Coeckelberg 2018: 150). Doch statt einer anthropozentrischen Sichtweise handelt es sich hier eher um eine relationale Sichtweise, die nichtmenschliche Wesen als Interaktionspartner behandelt.

## 4. Fazit

Künstliche Intelligenz im Allgemeinen und humanoide Roboter im Besonderen werden unser Leben und vielleicht auch uns selbst verändern. Philosophen müssen im Hinblick auf die epistemischen, ethischen, ästhetischen und politischen Auswirkungen dieser neuen Herausforderungen viel bedenken. Empathie ist nur eines der vielen Themen, die HRI in Frage stellt. Dieses Papier hat zu den notwendigen Untersuchungen beigetragen, die bereits laufen oder noch anstehen. Ich habe das epistemische Rätsel diskutiert, ob wir uns in Roboter einfühlen können, und habe dabei die vorherrschenden zeitgenössischen Darstellungen von Empathie auf diesen Bereich angewendet. Anschließend ging ich der normativen Frage nach, ob und warum wir uns in Roboter einfühlen sollten. Der Artikel schlug einen pragmatistischen Standpunkt vor, indem er zeigte, dass a) wir uns tatsächlich in humanoide Roboter einfühlen können, nicht nur auf einer grundlegenden Ebene, sondern zumindest bis zu einem gewissen Grad auch auf der Ebene der fantasievollen Perspektivenübernahme; Darüber hinaus wurde gezeigt, dass auch aus phänomenologischer und intersubjektiver Sicht von einer Empathie mit Robotern gesprochen werden kann, die in unsere Welt eingebettet sind, mit denen wir interagieren und eine kontextuelle Erzählung teilen. Der Fokus liegt auf Empathie als einem Prozess der gegenseitigen Interaktion und nicht als Ergebnis. Das Papier argumentierte jedoch auch, dass b) wir uns in humanoide Roboter einfühlen sollten, weil wir dadurch neues Wissen über ein sehr unbekanntes Wesen in der Welt erlangen und so unseren Horizont erweitern, für zukünftige KI-Entwicklungen trainieren und die HRI verbessern können ein gemeinsames soziales Umfeld. Dies sei nicht nur von instrumentellem Wert, sondern auch wertvoll für unser Verständnis von uns selbst und unserer Gesellschaft, in der Roboter und andere Formen der KI als Begle

[24] Coeckelbergh schlägt einen ähnlichen Ansatz wie ich vor, lässt sich jedoch von Wittgensteins Konzepten einer Lebensform und Sprachspielen inspirieren. Dennoch fehlt in seinem Aufsatz eine klare Definition dessen, was Empathie seiner Meinung nach bedeutet (z. B. ob Empathie tatsächlich die Sorge um das Wohlergehen des anderen beinhaltet, wie sein Aufsatz nahezulegen scheint).

## Verweise


Baron-Cohen, S. 1995. Geistesblindheit. Ein Essay über Autismus und Theorie des Geistes. Cambridge, MA: MIT Press.

Batson, CD 2009. Diese Dinge nennt man Empathie: Acht verwandte, aber unterschiedliche Phänomene. In „Die soziale Neurowissenschaft der Empathie“, hrsg. J. Decety und W. Ickes, 3–15. Cambridge, MA: MIT Press.

Benford, G. und E. Malartre. 2007. Jenseits des Menschen. Tom Doherty Associates: Leben mit Robotern und Cyborgs. New York.

Boddington, P., P. Millican und M. Wooldridge. 2017. Sonderausgabe „Minds and Machines“: Ethik und künstliche Intelligenz. Köpfe und Maschinen 27 (4): 569–574.

Boden, MA 2016. KI. Sein Wesen und seine Zukunft. Oxford: Oxford University Press.

Breazeal, CL 2002. Soziale Roboter entwerfen. Cambridge, MA: MIT Press.

Breithaupt, F. 2019. Die dunklen Seiten der Empathie. Ithaka: Cornell University Press.

Bretan, M., G. Hoffman und G. Weinberg. 2015. Emotional ausdrucksstarkes, dynamisches körperliches Verhalten bei Robotern. International Journal of Human-Computer Studies 78: 1–16.

Brinck, I. und C. Balkenius. 2018. Gegenseitige Anerkennung in der Mensch-Roboter-Interaktion: Eine deflationäre Darstellung. Philosophie und Technologie: 1–18. https://doi.org/10.1007/s13347-018-0339-x.

Chalmers, D. J. 1996. Das Bewusstsein. Oxford: Oxford University Press.

Coeckelberg, M. 2018. Warum sich für Roboter interessieren? Empathie, moralisches Ansehen und die Sprache des Leidens. Kairos. Zeitschrift für Philosophie und Wissenschaft 20: 141–158.

Colombetti, G. 2013. Der Gefühlskörper. Affektive Wissenschaft trifft auf den aktiven Geist. Cambridge, MA: MIT Press.

Coplan, A. 2011. Empathie verstehen, 3–18. Seine Eigenschaften und Wirkungen. In Empathie. Philosophische und psychologische Perspektiven. Oxford: Oxford University Press.

Coplan, A. und P. Goldie. 2011. Empathie. Philosophische und psychologische Perspektiven. Oxford: Oxford Universitätsverlag.

Cross, ES, Riddoch, KA, Pratts, J, Titone, S, Chaudhury, B und Hortensius, R. 2018. Eine neurokognitive Untersuchung der Auswirkungen der Geselligkeit mit einem Roboter auf die Empathie für Schmerzen. Vordruck. https://doi. org/10.1101/470534.

Darling, K. 2016. Ausweitung des Rechtsschutzes auf soziale Roboter: Die Auswirkungen von Anthropomorphismus, Empathie und gewalttätigem Verhalten gegenüber Roboterobjekten. In Robot law, hrsg. M. Froomkin, R. Calo und I. Kerr. Cheltenham: Edward Elgar.

Darwall, S. 1998. Empathie, Sympathie, Fürsorge. Philosophische Studien 89: 261–282.

De Sousa, R. 1987. Die Rationalität der Emotionen. Cambridge, MA: MIT Press.

De Vignemont, F. und P. Jacob. 2012. Wie ist es, den Schmerz eines anderen zu spüren? Wissenschaftstheorie 79 (2): 295–316.

De Vignemont, F. und T. Singer. 2006. Das empathische Gehirn: Wie, wann und warum? Trends im kognitiven Bereich Wissenschaften 10(10): 435–441.

Dennett, D. 1991. Bewusstsein erklärt. Boston: Little, Brown und Co.

Dullstein, M. 2012. Die zweite Person in der Debatte über die Theorie des Geistes. Rezension von Philosophie und Psychologie 3 (2): 231–248.

Dullstein, M. 2013. Direkte Wahrnehmung und Simulation: Steins Bericht über Empathie. Rezension von Philosophie und Psychologie 4: 333–350.

Dumouchel, P. und L. Damiano. 2017. Leben mit Robotern. Cambridge, MA: Harvard University Press.

Engelen, EM 2018. Können wir ein Wir-Gefühl mit einer digitalen Maschine teilen? Emotionales Teilen und das Anerkennung des einen als anderen. Interdisziplinäre wissenschaftliche Rezensionen 43 (2): 125–135.

Engelen, EM und B. Röttger-Rössler. 2012. Aktuelle disziplinäre und interdisziplinäre Debatten zum Thema Empathie. Emotionsrückblick 4 (1): 3–8.

Fodor, J. 1987. Psychosemantik. Das Bedeutungsproblem in der Philosophie des Geistes. Cambridge, MA: MIT Drücken Sie.

Gallagher, S. 2008. Direkte Wahrnehmung im interaktiven Kontext. Bewusstsein und Erkenntnis 17 (2): 535–543.

Gallagher, S. 2017. Empathie und Theorien der direkten Wahrnehmung. Im Routledge-Handbuch der Philosophie von Empathie, Hrsg. H. Maibom, 158–168. New York: Routledge.

Gallagher, S. und J. Gallagher. 2019. Sich wie ein anderer verhalten: Das Einfühlungsvermögen eines Schauspielers für seinen Charakter. Topoi (zuerst online), https://doi.org/https://doi.org/10.1007/s11245-018-96247.

Gallagher, S. und D. Hutto. 2008. Andere durch primäre Interaktion und Erzählpraxis verstehen. In The Shared Mind: Perspectives on Intersubjectivity, hrsg. J. Zlatev, T. Racine, C. Sinha und E. Itkonen, 17–38. Amsterdam/Philadelphia: John Benjamins Publishing Company.

Gallese, V. 2001. Die Hypothese der „gemeinsamen Mannigfaltigkeit“: Von Spiegelneuronen zu Empathie. Zeitschrift für Bewusstseinsstudien 8: 33–50.

Goldie, S. 2000. Die Emotionen. Oxford: Oxford University Press.

Goldie, P. 2012. Das Chaos im Inneren. Erzählung, Emotion und der Geist. Oxford: Oxford University Press.

Goldman, A. 2006. Simulation von Gedanken: Die Philosophie, Psychologie und Neurowissenschaft des Gedankenlesens. Oxford: Oxford University Press.

Goldman, A. 2011. Zwei Wege zur Empathie: Erkenntnisse aus der kognitiven Neurowissenschaft. In Empathie: Philosophische und psychologische Perspektiven, hrsg. A. Coplan und P. Goldie, 31–44. Oxford: Oxford University Press.

Gopnik, A. und H.M. Wellman. 1994. Die theoretische Theorie. In „Mapping the mind: Domain specifity in cognition and culture“, hrsg. L.A. Hirschfeld und S.A. Gelman, 257–293. Cambridge: Cambridge University Press.

Gruen, L. 2009. Sich um die Natur kümmern: Einfühlsamer Umgang mit der mehr als menschlichen Welt. Ethik und die Umwelt 14 (2): 23–38.

Gruen, L. 2017. Der moralische Status von Tieren. In der Stanford Encyclopedia of Philosophy (Herbstausgabe 2017) Hrsg. DE Zalta, https://plato.stanford.edu/archives/fall2017/entries/moral-animal/.

Hickok, G. 2014. Der Mythos der Spiegelneuronen: Die wahre Neurowissenschaft von Kommunikation und Kognition. Neu York: WW Norton & Company.

Hoffmann, M. und R. Pfeifer. 2018. Roboter als mächtige Verbündete für die Untersuchung verkörperter Kognition von Grund auf. Im Oxford Handbook of 4E Cognition, hrsg. A. Newen, L. de Bruin und S. Gallagher. Oxford: Oxford University Press.

Hutto, DD 2008. Die Narrativpraxis-Hypothese: Klarstellungen und Implikationen. Philosophische Erkundungen 11 (3): 175–192.

Iacoboni, M. 2011. Ineinander: Neuronale Mechanismen für Empathie im Primatengehirn. In Empathie: Philosophische und psychologische Perspektiven, hrsg. A. Coplan und P. Goldie, 45–57. Oxford: Oxford University Press.

Iacoboni, M., R. P. Woods, et al. 1999. Kortikale Mechanismen der menschlichen Nachahmung. Wissenschaft 286: 2526–2528.

Kanske, P. 2018. Der soziale Geist: affektive und kognitive Methoden entwirren, um andere zu verstehen. Interdisziplinäre wissenschaftliche Rezensionen 43 (2): 115–124.

Kant, I. 1997. Vorlesungen zur Ethik, hrsg. und trans. P. Heath und JB Schneewind. Cambridge: Cambridge Universitätsverlag.

Kasparov, G. 2017. Tiefgründiges Denken: Wo maschinelle Intelligenz endet und menschliche Kreativität beginnt. New York: Öffentliche Angelegenheiten.

Leite, A., A. Pereira, S. Mascarenhas, C. Martinho, R. Prada und A. Paiva. 2013. Der Einfluss von Empathie auf die Mensch-Roboter-Beziehungen. International Journal of Human-Computer Studies 71(3): 250–260.

Lin, P., R. Jenkins und K. Abney. 2017. Roboterethik 2.0: Von autonomen Autos bis hin zu künstlicher Intelligenz. Oxford: Oxford University Press.

Loh, J. 2019. Roboterethik. Eine Einführung. Berlin: Suhrkamp.

MacLennan, BJ 2014. Ethischer Umgang mit Robotern und das schwierige Problem der Roboteremotionen. International Journal of Synthetic Emotions 5 (1): 9–16.

Maibom, H. 2017. Das Routledge-Handbuch zur Philosophie der Empathie. London: Routledge.

Misselhorn, C. 2009. Empathie mit unbelebten Objekten und dem unheimlichen Tal. Köpfe und Maschinen 19 (3): 345–359.

Misselhorn, C. Im Druck. Ist Empathie mit Robotern moralisch relevant? In Emotional Machines: Perspectives from Affective Computing and Emotional Human-Machine Interaction, hrsg. C. Misselhorn und M. Klein. Wiesbaden.

Mori, M. 2005. Über das unheimliche Tal. In Proceedings of the Humanoids-2005 Workshop: Ansichten des unheimlichen Tals. Tsukuba: Japan.

Nagel, T. 1974. Wie ist es, eine Fledermaus zu sein? Die philosophische Rezension 83(4): 435–450.

Newen, A. 2015. Andere verstehen: Die Personenmodelltheorie. In Open MIND: 26(T), hrsg. T. Metzinger und JM Windt. Frankfurt am Main: MIND Group.

Newen, A., L. De Bruin und S. Gallagher. 2018. Das Oxford-Handbuch zur 4E-Kognition. Oxford: Oxford Universitätsverlag.

Nussbaum, M. 2011. Umbrüche des Denkens: Die Intelligenz der Emotionen. Cambridge: Universität Cambridge Drücken Sie.

Plantinga, C. 2009. Zuschauer bewegen: Amerikanischer Film und die Erfahrung des Zuschauers . Berkeley: Universität Kalifornische Presse.

Rorty, R. 2001. Erlösung vom Egoismus: James und Proust als spirituelle Übungen. Telos 3 (3): 243–263.

Scheutz, M. 2011. Architektonische Affektrollen und wie man sie in künstlichen Agenten bewertet. International Journal of Synthetic Emotions 2(2): 48–65.

**Schmetkamp, S. 2017. Perspektiven auf unser Leben gewinnen: Stimmungen und ästhetische Erfahrung. Philosophie 45(4): 1681–1695.**

**Schmetkamp, S. 2019. Theorien der Empathie – Eine Einführung. Hamburg: Junius Verlag.**

**Schneider, S. Im Druck. Zukünftige Köpfe: Das Gehirn verbessern und transzendieren.**

**Slote, M. 2017. Die vielen Gesichter der Empathie. Philosophia 45 (3): 843–855.**

**Smith, M. 1995. Fesselnde Charaktere: Fiktion, Emotionen und das Kino. Oxford: Clarendon Press.**

**Sobchack, V. 2004. Fleischliche Gedanken: Verkörperung und Bewegtbildkultur. Berkeley: Universität Kalifornische Presse.**

**Stein, E. 1989. Zum Problem der Empathie: Die gesammelten Werke von Edith Stein. Bd. 3 (3. überarbeitete Auflage), trans. W. Stein. Washington, DC: ICS-Veröffentlichungen.**

**Stueber, K. 2006. Empathie wiederentdecken: Handlungsfähigkeit, Volkspsychologie und Geisteswissenschaften. Cambridge, MA: MIT Press.**

**Stüber, K. 2018. Empathie. In der Stanford Encyclopedia of Philosophy (Ausgabe Frühjahr 2018), hrsg. DE Zalta, https://plato.stanford.edu/archives/spr2018/entries/empathy/.**

**Vaage, MB 2010. Spielfilme und die Spielarten empathischen Engagements. Midwest Studies in Philosophie 34: 158–179.**

**Vallor, S. 2011. Carebots und Pflegekräfte: Aufrechterhaltung des ethischen Ideals der Pflege im 21. Jahrhundert. Philosophie und Technologie 24 (3): 251–268.**

**Weber, K. 2013. Wie ist es, einem autonomen künstlichen Agenten zu begegnen? KI & GESELLSCHAFT 28: 483–489.**

**Yanal, RJ 1999. Paradoxien von Emotion und Fiktion. Pennsylvania: Penn State University Press.**

**Zahavi, D. 2011. Empathie und direkte soziale Wahrnehmung: Ein phänomenologischer Vorschlag. Rezension von Philosophie und Psychologie 2 (3): 541–558.**

**Zahavi, D. 2014. Selbst und andere: Erforschung von Subjektivität, Empathie und Scham. Oxford: Universität Oxford Drücken Sie.**

**Zahavi, D. und J. Michael. 2018. Jenseits des Spiegelns: 4E-Perspektiven zur Empathie. Im Oxford Handbook of 4E Cognition, hrsg. A. Newen, L. de Bruin und S. Gallagher, 589–606. Oxford: Oxford University Press.**