
Mara, M., & Leichtmann, B. (in press). Soziale Robotik und Roboterpsychologie: Wie psychologische Forschung zur menschenzentrierten Entwicklung robotischer Systeme beitragen kann. In O. Bendel (Ed.), *Soziale Roboter*. Springer.

MANUSKRIFT DER AUTOR*INNEN

9 Soziale Robotik und Roboterpsychologie

Wie psychologische Forschung zur menschenzentrierten Entwicklung robotischer Systeme beitragen kann

Martina Mara und Benedikt Leichtmann

Zusammenfassung

Egal in welchem Anwendungsbereich soziale Roboter zum Einsatz kommen, sie sind in jedem Fall dafür gemacht, mit Menschen zu interagieren. Die Psychologie als Wissenschaft menschlichen Denkens, Fühlens und Verhaltens kann im Austausch mit anderen Disziplinen daher wertvolle Beiträge zur Erforschung sozialer Roboter und zum Gelingen von Mensch-Roboter-Interaktion liefern. Ziel dieses Kapitels ist es, einen einführenden Einblick in die Besonderheiten psychologischer Zugänge zur Sozialen Robotik zu geben. Ausgewählte Themenfelder der Roboterpsychologie, darunter Anthropomorphismus, Technologieängstlichkeit oder Vertrauen in Roboter, werden theoretisch und empirisch beleuchtet. Da eine korrekte Einordnung von Forschungsergebnissen immer auch ein Verständnis zugrunde liegender Forschungsmethoden erfordert, geht ein weiterer Abschnitt darauf ein, wie psychologische Vorgänge wissenschaftlich untersucht und gemessen werden können. Mit der Frage „Brauchen Roboter Psychotherapie?“ widmen wir uns abschließend populären Falschvorstellungen zur Roboterpsychologie und diskutieren außerdem Potenziale und Risiken von Robotern als therapeutische Begleitwerkzeuge.

„Wo soziale Roboter sind, sind Menschen. Und wo Menschen sind, ist die Psychologie.“

9.1 Zur Relevanz der Psychologie für die Robotik

Soziale Roboter können positive Auswirkungen auf Mensch und Gesellschaft haben. Konzepte zum Praxiseinsatz sozialer Roboter sehen beispielsweise vor, dass diese beim Lernen unterstützen oder kleine Aufgaben im Alltag erledigen könnten, dass sie zum Erhalt von Autonomie und zur Reduktion von Einsamkeit bei älteren Personen beitragen könnten oder dass sie einfach nur Unterhaltung bieten. Werden die Bedürfnisse, Wahrnehmungen und Ziele von menschlichen Nutzer*innen bei der Entwicklung und Gestaltung neuer Roboter aber nicht ausreichend berücksichtigt, kann dies schnell zu nachteiligen Konsequenzen führen. Beispiele hierfür wären eine geringe Akzeptanz und damit einhergehende Ablehnung von Robotern, Fehler in deren Benutzung, Unfälle oder ungewollte psychische Effekte wie etwa ein Gefühl des Dominiertwerdens durch Technologie.

Die Einführung neuer Roboter in bestehende soziale Systeme setzt daher immer einen stark menschenzentrierten Entwicklungsansatz voraus, bei dem die intendierten Nutzer*innen und ihre oftmals unterschiedlichen Bedürfnisse und Erlebenswelten im Zentrum stehen. Mit ihren Theorien und Methoden zur Erforschung menschlichen Denkens, Fühlens und Verhaltens liefert die Psychologie in diesem Zusammenhang wertvolle Beiträge und ist im Zusammenspiel mit anderen Disziplinen wesentlich an der Untersuchung und Implementierung sozialer Roboter beteiligt.

9.1.1 Was ist eigentlich Psychologie?

Dort, wo Menschen sind, ist auch die Psychologie. Ganz allgemein wird die Psychologie als wissenschaftliche Disziplin definiert, die sich mit der Untersuchung menschlichen Erlebens und Verhaltens beschäftigt. Genauso vielfältig wie die Psyche des Menschen sind auch die Anwendungsgebiete der Psychologie. Sie gehen über klassische Felder wie die klinisch-psychologische Diagnostik und die Psychotherapie, die vielen wahrscheinlich als erstes in den Sinn kommen, weit hinaus und reichen überall dort hinein, wo Menschen miteinander oder mit ihrer Umwelt interagieren. Subdisziplinen wie die Arbeits- und Organisationspsychologie, die Rechtspsychologie, die Verkehrspsychologie oder gar die Raumfahrtpsychologie geben einen kleinen Einblick in die Breite

psychologischen Wirkens.

Ein mit zunehmender Technologisierung und Digitalisierung immer wichtiger werdendes Anwendungsgebiet ist auch die Mensch-Technik-Interaktion, wodurch historisch gesehen jüngere Gebiete der Angewandten Psychologie wie beispielsweise die Medien- und Kommunikationspsychologie oder die Ingenieurpsychologie starken Aufschwung erfuhren. In diesen Teildisziplinen untersuchen Psycholog*innen die technikvermittelte Interaktion von Menschen untereinander oder die direkte Interaktion zwischen Menschen und Technik und bedienen sich dabei psychologischer Theorien, Methoden und Forschungspraktiken. Dabei übernimmt die Psychologie nicht nur rein evaluatorische Aufgaben im Sinne eines nachträglichen Bewertenlassens bereits existenter Produkte, sondern ist bestenfalls Bestandteil eines interdisziplinären Gesamtprozesses, der von der Berücksichtigung bestehender Forschungsergebnisse in ersten konzeptuellen Überlegungen und Entwicklungsschritten bis hin zur Implementierung in reale Kontexte reicht.

Zudem verfolgen Psycholog*innen auch eigene genuin psychologische Fragestellungen und Forschungsziele. Beispiele für den Bereich der Robotik wären: Wie nehmen Menschen robotische Technologien wahr – als technische Werkzeuge oder als eigenständige soziale Akteure? Wie können komplexe psychologische Konstrukte, wie die Einstellung gegenüber Robotern, verlässlich gemessen werden? Wie wirkt sich die Einführung einer neuen Robotertechnologie in einem Unternehmen auf die Arbeitszufriedenheit der Mitarbeiter*innen aus? Wie entwickelt sich das Vertrauen von Nutzer*innen in einen sozialen Roboter über die Zeit und aufgrund welcher Parameter unterscheiden sich verschiedene Nutzer*innen in ihrer Vertrauensbildung? Die Untersuchung der Mensch-Roboter-Interaktion nach psychologischen Gesichtspunkten ist dadurch nicht nur für die menschenfreundliche Gestaltung ebenjener hilfreich, sondern stellt für die Psychologie auch eine weitere Möglichkeit dar, psychologische Theorien, Methoden und Interventionen in neuen, spannenden und zunehmend relevanten Anwendungsszenarien zu validieren (siehe z.B. Echterhoff et al. 2006).

Als noch neues Feld der anwendungsorientierten psychologischen Forschung beschäftigt sich die Roboterpsychologie als Nachbarbereich der Ingenieurpsychologie, der Medienpsychologie und der Mensch-Computer-Interaktion mit der Erklärung, Beschreibung und Veränderung davon, wie unterschiedliche Menschen (teil-)autonome maschinelle Agenten erleben, sowie ihren Einstellungen und Verhaltensweisen gegenüber diesen. Maschinelle Agenten können hierbei sowohl physisch verkörperte Roboter (z.B. Assistenzroboter) als auch nichtverkörperte Systeme (z.B. Chatbots) sein. Im Gegensatz etwa zu soziologischen Betrachtungen der Robotik nimmt die Roboterpsychologie in diesem Kontext primär Prozesse und Einflussfaktoren auf Ebene des Individuums in den Fokus.

9.1.2 Das Individuum in Interaktion mit sozialen Robotern

Die Psychologie, und als eine ihrer Subdisziplinen auch die Roboterpsychologie, untersucht also die mentalen Prozesse und das Verhalten von einzelnen Menschen, von Individuen (vgl. Gerrig 2016). Psycholog*innen gehen davon aus, dass individuelles Erleben und Verhalten nicht allein entweder durch Merkmale des Individuums oder durch Merkmale der Situation, in der sich das Individuum befindet, erklärt werden kann, sondern am besten durch das Zusammenwirken beider. Wie ängstlich sich eine bestimmte Person gegenüber einem sozialen Roboter zeigt, wie sehr sie ihm menschliche Eigenschaften zuspricht oder wie viel Vertrauen sie ihm entgegenbringt, hängt dieser Annahme zufolge daher nicht nur von Eigenschaften des Roboters (z.B. Aussehen, Verhalten) und dem Kontext des Aufeinandertreffens (z.B. Pflegekontext, Unterhaltungskontext), sondern auch von Eigenschaften, die bei der Person selbst liegen, ab.

Eigenschaften der Person, die im Rahmen psychologischer Forschung zu sozialen Robotern typischerweise von Interesse sein können, sind soziodemografische Merkmale (z.B. Alter, Bildung, Familienstand), Erfahrung (z.B. Routine im Umgang mit einem bestimmten Roboter), Einstellungsfaktoren (z.B. allgemeine Einstellung gegenüber neuen Technologien) oder Persönlichkeitsfaktoren. Im Gegensatz zu aktuellen Stimmungen und Zuständen einer Person (engl. *states*) werden Persönlichkeitsfaktoren (engl. *traits*) als relativ stabile, zeitlich überdauernde Eigenschaften von Individuen betrachtet. Zu den bekanntesten psychologischen Modellen zur Beschreibung interindividueller Persönlichkeitsunterschiede gehört das Fünf-Faktoren-Modell (auch Big Five oder OCEAN-Modell genannt) (John und Srivastava 1999). Diesem Modell zufolge lässt sich

jeder Mensch entlang der folgenden fünf Hauptdimensionen der Persönlichkeit einordnen: 1) Offenheit (Wie neugierig und offen für neue Erfahrungen ist jemand?), 2) Gewissenhaftigkeit (Wie zuverlässig und organisiert ist jemand?), 3) Extraversion (Wie gesprächig und aktiv ist jemand?), 4) Verträglichkeit (Wie freundlich und kooperativ ist jemand?) und 5) Neurotizismus (Wie ängstlich und nervös ist jemand?; auch emotionale Instabilität genannt).

Der Einfluss dieser fünf Persönlichkeitsmerkmale wurde bereits in einigen Studien im Bereich der Mensch-Roboter-Interaktion betrachtet. In einer Metaanalyse von Esterwood und Kolleg*innen zeigte sich auf Basis von 26 inkludierten Studien etwa, dass höhere Werte bei Offenheit, Extraversion und Verträglichkeit in Zusammenhang mit einer höheren Akzeptanz von Robotern stehen (Esterwood et al. 2021). Für extrovertierte Personen liegen beispielsweise auch Hinweise dafür vor, dass diese im Vergleich zu introvertierten Personen Roboter stärker vermenschlichen (Salem et al. 2015). Emotional instabilere Individuen und solche, die zu Perfektionismus neigen, scheinen hingegen menschenähnliche Roboter als besonders unheimlich zu empfinden (MacDorman und Entezari 2015).

Neben den populären Big Five lassen sich aus der Persönlichkeitspsychologie noch viele weitere Charakteristika ableiten, die in Zusammenhang mit Effekten sozialer Roboter oder mit deren Akzeptanz stehen könnten. Beispielhaft seien an dieser Stelle Ambiguitätstoleranz und die individuelle Vertrauenstendenz als potenziell interessante Variablen genannt. Personen, die mehrdeutige Situationen oder widersprüchliche Handlungsweisen insgesamt als unangenehm empfinden und dementsprechend eine niedrige Ambiguitätstoleranz aufweisen, könnten auch Roboter, die zugleich menschlich und maschinenhaft wirken, eher ablehnen (vgl. Lischetzke et al. 2017). Für Individuen, die in ihrem Leben generell stärker dazu neigen, anderen Personen Vertrauen entgegenzubringen, ist die Annahme naheliegend, dass diese auch Robotern eine höhere Vertrauenswürdigkeit zuschreiben würden (vgl. Koerber 2018).

9.2 Ausgewählte psychologische Theoriefelder

Es existieren bereits viele psychologische Konstrukte, Theorien und Forschungsbefunde zur Beschreibung und Vorhersage menschlichen Denkens, Fühlens und Verhaltens, die sich auf den neuen Kontext der Interaktion mit sozialen Robotern übertragen lassen. Nichtsdestotrotz unterscheiden sich Anwendungskontexte natürlich untereinander, weshalb Theorien angepasst oder um zusätzlich relevante Einflussfaktoren erweitert werden müssen. In der Sozialen Robotik wird dabei häufig auf Erkenntnisse aus der Sozialpsychologie, in deren Zentrum üblicherweise zwischenmenschliche Interaktionen stehen, zurückgegriffen (Echterhoff et al. 2006; Reeves und Nass 1996). In Anwendung auf Mensch-Maschine-Beziehungen können sozialpsychologische Theorien dann zum Beispiel um den Einfluss von Anthropomorphismus, d.h. um die Zuschreibung menschlicher Eigenschaften gegenüber Maschinen oder anderen nichtmenschlichen Entitäten (Epley et al. 2007), ergänzt und neu überprüft werden.

Im folgenden Abschnitt werden einige ausgewählte Themenfelder, die in der Roboterpsychologie von Relevanz sind, vorgestellt. Beachtet werden sollte, dass diese Auswahl keinesfalls einen vollumfänglichen Überblick verkörpert. Stattdessen sollen einführende Einblicke in interessante psychologische Fragestellungen der Mensch-Roboter-Interaktionen zur weiteren eigenständigen Recherche und Auseinandersetzung inspirieren.

9.2.1 Soziale Perzeption und Anthropomorphismus

Wie nehmen Menschen robotische Technologien wahr – als Maschinen und Werkzeuge, als soziale Akteure mit menschenähnlichen Charakteristika, oder als etwas dazwischen? Unter welchen Bedingungen werden Robotern (unbewusst) soziale Eigenschaften zugesprochen? Solche und ähnlich gelagerte Fragen zur sozialen Perzeption von Robotern werden häufig wissenschaftlich untersucht.

Basierend auf klassischen Theorien und Experimenten aus der Sozialpsychologie stellten Reeves und Nass (1996) bereits Mitte der 1990er-Jahre ihre Hypothese zur „Media Equation“ oder auch zum CASA-Paradigma („Computers are Social Actors“) auf. Diese besagt, dass Menschen gegenüber technischen Geräten die gleichen sozialen Verhaltensweisen zeigen wie gegenüber menschlichen

Interaktionspartner*innen. Nach dieser Hypothese würden Menschen beispielsweise ihre Antworten gegenüber einem Computer aus Höflichkeit in eine sozial erwünschte Richtung verzerren, anstatt dem Gerät ehrliches Feedback über die Nützlichkeit seiner Funktionen zu geben (Nass et al. 1999). Neuere Arbeiten zeigen jedoch, dass es hier einer differenzierteren Betrachtung bedarf. So konnten Leichtmann und Nitsch (2020b) in einer konzeptionellen Replikation der angesprochenen Studie von Nass et al. (1999), in der Roboter statt Computer verwendet wurden, kein solches Höflichkeitsverhalten feststellen. Auch in anderen Studien blieb ein solcher Effekt oft aus (siehe Leichtmann und Nitsch 2020b). Das könnte vermuten lassen, dass Technologien entgegen der CASA-Hypothese nicht immer automatisch als soziale Akteure betrachtet werden. Wie stark Menschen Robotern soziale Eigenschaften zuschreiben, variiert demnach in Abhängigkeit von Gestaltung der Technologie, Eigenschaften der Nutzer*innen oder der Situation. Es ist also nicht die Frage, ob die CASA-Hypothese grundsätzlich korrekt ist, sondern vielmehr unter welchen Bedingungen Menschen welcher Art von Roboter wie stark welche sozialen Eigenschaften zuschreiben.

Ein wichtiges Konstrukt in diesem Zusammenhang ist Anthropomorphismus. Während es keine exakte einheitliche Definition gibt, lässt sich Anthropomorphismus grundsätzlich als die Tendenz beschreiben, nichtmenschlichen Entitäten (z.B. einem Baum, Haustier, Auto oder Computer) menschenähnliche Eigenschaften, Intentionen oder Gefühlszustände zuzuschreiben (Epley et al. 2007) – und somit natürlich auch Robotern. Arbeiten von Waytz und Kolleg*innen (2010) zeigen, dass sich Menschen grundsätzlich in der Stärke ihrer persönlichen Tendenz zum Anthropomorphismus unterscheiden. Zudem hängt es auch von der Situation ab, wie stark wir anthropomorphisieren. So besagt beispielsweise die Drei-Faktoren-Theorie des Anthropomorphismus (Epley et al. 2007), dass wir vor allem dann Menschliches im Nichtmenschlichen sehen, wenn 1) Wissen um menschliche Eigenschaften als Basis für Anthropomorphismus zugänglich und anwendbar ist, 2) uns Anthropomorphismus als effektive Strategie dient, um nichtmenschliche Agenten besser einordnen zu können und ihr Verhalten vorherzusagen, und 3) wenn das Verlangen nach menschlichem Kontakt besonders hoch ist, was etwa in Zeiten sozialer Isolation der Fall sein könnte. Psychologische Untersuchungen zu Anthropomorphismus im Bereich der Sozialen Robotik zeigen, dass diese Theorie auch zur Erklärung von Unterschieden in der Wahrnehmung von Robotern herangezogen werden kann. So geht beispielsweise aus einer Laborstudie von Eyssel und Reich (2013) hervor, dass Personen, bei denen zuvor das Gefühl von Einsamkeit induziert wurde und die demnach einen größeren Wunsch nach menschlichem Kontakt haben sollten, einem Roboter stärker menschliche Eigenschaften (z.B. einen eigenen Verstand) zuschrieben als Personen mit neutraler Stimmungslage.

Wie anthropomorph ein sozialer Roboter wahrgenommen wird, kann darüber hinaus auch von anderen Faktoren abhängen. Dass der Grad seiner optischen Menschenähnlichkeit einen zentralen Einflussfaktor darstellt, mag wenig überraschen. Unabhängig vom Äußeren kann aber auch das nonverbale Verhalten des Roboters (Mara und Appel 2015a) oder die Art, wie der Roboter seinen Interaktionspartner*innen vorgestellt wird, auf anthropomorphe Zuschreibungen einwirken. In einem Feldexperiment in einem Technologiemuseum nahmen Besucher*innen einen Roboter zum Beispiel dann als besonders menschenähnlich wahr, wenn dieser durch eine fiktionale Kurzgeschichte anstatt eines üblichen nichtnarrativen Informationstextes eingeführt wurde (Mara und Appel 2015b).

9.2.2 Aversive Reaktionen auf menschenähnliche Roboter

Ist Menschenähnlichkeit bei sozialen Robotern stets mit positiven Evaluierungen und angenehmen Erlebnissen verbunden? Vor dem Hintergrund der sogenannten „Uncanny Valley“-Hypothese (dtsh. „unheimliches Tal“), die der japanische Robotiker Masahiro Mori im Jahr 1970 aufgestellt hat und die in jüngeren Jahren auch Gegenstand psychologischer Forschungsarbeiten geworden ist, ist das nicht der Fall. Die Hypothese beschreibt einen kurvilinearen Zusammenhang zwischen dem Grad der Menschenähnlichkeit einer künstlichen Figur und den Reaktionen menschlicher Beobachter*innen oder Nutzer*innen: Solange die Figur – stellen wir uns einen sozialen Roboter vor – eindeutig als Maschine erkennbar ist, geht anwachsende Menschenähnlichkeit auch mit zunehmend positiven Bewertungen einher. Dieser Trend sollte sich aber umkehren, sobald der Roboter ein Level an sehr hohem, wenngleich noch nicht perfektem, menschlichen Realismus erreicht. In diesem Bereich, in dem beispielsweise lebensgroße Androide mit Silikonhaut und animierter Mimik vermutet werden könnten, sinkt die Akzeptanz ab und die Figur sollte als unheimlich, gruselig oder gar bedrohlich erlebt

werden. Erst perfekte Kopien des Menschen würden Mori zufolge den Kurvenverlauf abermals revidieren und rechts des „unheimlichen Tals“ wiederum zu steigender Akzeptanz führen (Mori 1970).

Obwohl einige Forschungsarbeiten empirische Unterstützung für die zentralen Annahmen Moris fanden (z.B. Burleigh et al. 2013; Mathur und Reichling 2016), zeigten sich bei anderen Autor*innen keine Hinweise auf nichtlineare Unheimlichkeitseffekte (z.B. Bartneck et al. 2009; Kätsyri et al. 2015). Angesichts der Tatsache, dass bisher nur wenige androide Roboter mit ausreichend hohem menschlichen Realismusgrad existieren und viele Studien zur Untersuchung des Uncanny Valley daher auf weniger menschlich wirkende Roboter oder auf Foto- bzw. Videostimuli zurückgreifen mussten, kann die Hypothese bisher weder eindeutig akzeptiert noch verworfen werden. Der Forschungsstand lässt hier also noch keine klaren Ableitungen zu, wird in den nächsten Jahren aber bestimmt fortlaufend erweitert werden.

Neben der Untersuchung aversiver Reaktionen auf die optische Menschenähnlichkeit von Robotern werden zunehmend auch Effekte verhaltensspezifischer Menschenähnlichkeit erforscht. Bisher vorliegende Arbeiten deuten darauf hin, dass Roboter, die hochgradig selbständig entscheiden und handeln können, und insbesondere auch solche, die emotionale Empfindsamkeit vortäuschen, unabhängig von ihrem äußeren Erscheinungsbild als unheimlicher wahrgenommen werden als Roboter, die als vom Menschen kontrollierbare Werkzeuge präsentiert werden (Appel et al. 2020, 2016). Als möglichen Mechanismus hinter einem solchen „Uncanny Valley of Mind“ nennen Stein und Ohler (2017) eine von Studienteilnehmer*innen wahrgenommene Bedrohung der Einzigartigkeit des Menschen, die mit der Darstellung technischer Systeme als denkende und fühlende Wesen einhergehen kann.

9.2.3 Vertrauensbildung in der Mensch-Roboter-Interaktion

Vertrauen ist ein wichtiges Fundament jeglicher Beziehung, eine Art sozialer Kitt. Kann man sich auf Personen, mit denen man im persönlichen oder beruflichen Umfeld häufig zu tun hat, nicht verlassen, geht das mit negativen Effekten auf das eigene Wohlbefinden einher. Auch in der Sozialen Robotik stellt sich die Frage, welche Eigenschaften Roboter mitbringen müssen, damit sie möglichst vertrauenserweckend und damit positiv auf ihre Interaktionspartner*innen wirken. Um sich dieser Frage widmen zu können, lohnt sich zuerst ein Blick darauf, was mit dem gebräuchlichen Begriff des Vertrauens in der Psychologie eigentlich gemeint ist. In der Literatur zu interpersonellem Vertrauen wurde Vertrauen definiert als „die Erwartung eines Individuums oder einer Gruppe, dass man sich auf das Wort, das Versprechen, die mündliche oder schriftliche Aussage eines anderen Individuums oder einer anderen Gruppe verlassen kann“ (Rotter 1967, S. 1). Im Grunde geht es also um erwartungskonformes (und unschädliches) Verhalten, was auch die American Psychological Association – die weltweit größte Vereinigung von Psycholog*innen – unterstreicht, wenn sie die Vorhersehbarkeit eines Interaktionspartners als Schlüsselfaktor für Vertrauen beschreibt (APA 2021).

Obwohl das Feld der Vertrauensforschung sehr weit ist, ist vielen theoretischen Ansätzen die Unterscheidung zweier grundlegender Routen der Vertrauensbildung gemein. Die erste Dimension bezieht sich auf affektive und soziale Grundfesten des Vertrauens. Demnach entsteht Vertrauen dann, wenn das Gegenüber als wohlgesinnt, warmherzig und gemeinsame moralische Werte befolgend wahrgenommen wird. Die zweite Dimension bezieht sich auf kognitive Aspekte. Demnach wirkt es sich positiv auf die Vertrauensbildung aus, wenn das Gegenüber als kompetent und zuverlässig im Sinne zu erbringender Leistungen wahrgenommen wird (z.B. McAllister 1995; Rempel et al. 1985). Während affektiv-soziales Vertrauen bei Technologien wie Industrierobotern oder autonomen Fahrzeugen eine untergeordnete Rolle spielt und dort der Fokus mehr auf Kompetenzwahrnehmung oder das Vorhersehbarmachen von Intentionen gelegt wird (vgl. Koerber 2018; Mara et al. 2021), könnte das bei sozialen Robotern, die in der Regel stärkere soziale Perzeptionen hervorrufen, anders sein.

Bisher vorliegende empirische Untersuchungen zu Voraussetzungen und Effekten von Vertrauen in soziale Roboter weisen allerdings noch eine große methodische Heterogenität auf (z.B. in der Art, wie Vertrauen gemessen wurde) und sind deshalb schwer miteinander vergleichbar. Eine Metaanalyse von Naneva und Kolleg*innen (2020) betrachtet aber immerhin 30 Studien, in denen Vertrauen in soziale Roboter untersucht wurde, und kommt zum Schluss, dass der über alle Studien hinweg

gewichtete Mittelwert für Vertrauen nahe bei Null lag, was darauf hindeutet, dass Menschen sozialen Robotern bisher weder besonders vertrauen noch misstrauen. Bei sozialen Robotern, für die ein Einsatz im Gesundheitswesen geschildert wurde, waren die Vertrauenswerte am niedrigsten. Eine zweite Metaanalyse von Stower et al. (2021) analysierte 20 Studien zu Vertrauen von Kindern in soziale Roboter. Einer der wenigen Zusammenhänge, der sich über die verschiedenen Untersuchungen hinweg statistisch festmachen ließ, war, dass Kinder humanoiden Robotern mit größerer visueller oder verhaltensspezifischer Menschenähnlichkeit im Mittel weniger Kompetenz und daher auch weniger kognitives Vertrauen zusprechen als Robotern, die nicht menschenähnlich sind. Begründet sein könnte dieser Zusammenhang in den größeren Erwartungen an die Leistung menschenähnlich wirkender Roboter, die bei Nichterfüllung potenziell zu einer Verringerung der Kompetenzwahrnehmung führen.

Worin sich viele Forscher*innen mittlerweile einig sind, ist, dass der Beitrag der Wissenschaft zu Vertrauen in der Mensch-Roboter-Interaktion nicht darin liegen kann, empirische Grundlagen für eine Steigerung der wahrgenommenen Vertrauenswürdigkeit von Robotern unter allen Umständen zu liefern, sondern dass eine differenziertere Betrachtung nötig ist. Ein „Zuviel“ an Vertrauen (engl. *overtrust*) kann nämlich auch Risiken mit sich bringen, beispielsweise wenn ein besonders harmlos und niedlich wirkender Roboter zum Einsatz kommt, um jemandem sensible Daten zu entlocken (Wolfert et al. 2020). Anstelle einer Akzeptanzoptimierung rücken daher zunehmend Fragen rund um eine angemessene, prozesshafte „Kalibrierung“ von Vertrauen (engl. *trust calibration*) in den Mittelpunkt des wissenschaftlichen Interesses (De Visser et al. 2020).

9.2.4 Nähe und Distanz in der physischen Interaktion mit Robotern

Ein zentraler Aspekt bei der Interaktion zwischen Menschen und sozialen Robotern ist, dass sich beide Akteure häufig ein und denselben Raum teilen. So bewegen sich menschliche Nutzer*innen durch die physischen Umgebungen, in denen Roboter agieren, und auch Roboter werden zunehmend mobiler und flexibler in ihrem Bewegungsradius. Wenn ein Roboter einem Menschen dabei zu schnell zu nahe kommt, kann das zu Unwohlsein und schließlich zu Rückzugs- und Vermeidungsverhalten seitens menschlicher Nutzer*innen führen (Aiello 1987). Ein zentrales Konstrukt, das daher bei der Gestaltung von Mensch-Roboter-Interaktion berücksichtigt werden sollte – beispielsweise im Bewegungsdesign von Robotern oder der Gestaltung von Arbeits- und Lebensräumen mit Robotern – ist der „Personal Space“ von Menschen (dtsch. „persönlicher Distanzraum“).

Gifford (2013) definiert Personal Space als „die dynamische räumliche Komponente zwischenmenschlicher Beziehungen [...], die sich in der wechselnden Distanz und dem Orientierungswinkel zwischen Individuen während ihrer Interaktion darstellt“ (Gifford 2013, S. 125, übersetzt aus dem Englischen). Ein frühes Modell zum menschlichen Distanzverhalten, bekannt vor allem unter dem Begriff „Proxemik“ (Hall 1966), unterscheidet dabei verschiedene Distanzzonen in Abhängigkeit von der zwischenmenschlichen Beziehungskonstellation. Solche Kategorisierungen von Personal Space wurden jedoch häufig kritisiert und verworfen, da eine starre Einteilung in Zonen der Veränderlichkeit von Wohlfühlabständen nicht genug Rechnung trägt und daher nicht ausreichend zur adäquaten Beschreibung von Distanzverhalten beitragen kann. In modernen Modellen zur Erklärung menschlichen Distanzverhaltens wird Distanz daher als dynamisches Kontinuum betrachtet, dessen Ausprägung von verschiedenen interpersonellen und situationalen Faktoren abhängt.

In einer Metastudie von Leichtmann und Nitsch (2020a) wurden bisherige Erkenntnisse zum Einfluss verschiedener Faktoren auf die Wohlfühlidistanz speziell im Kontext der Mensch-Roboter-Interaktion zusammengefasst. Das weite Spektrum an mittleren Wohlfühlabständen (zwischen 25 und 178 cm) zeigt deutlich, dass sich Ergebnisse zu präferierten Distanzen zu Robotern stark zwischen einzelnen Studien unterscheiden und daher kontextabhängig sein dürften. Obwohl bereits einige Studien existieren, ist noch relativ unklar, welche Faktoren in welcher Konstellation die stark unterschiedlichen Resultate erklären. Es wird auch vermutet, dass die Relevanz von Geschlechterunterschieden oder von Persönlichkeitsfaktoren in Bezug auf Nähe- und Distanzwahrnehmungen in der vergangenen Literatur überschätzt wurde. Lediglich der Einfluss der Erfahrung mit Robotern zeigte im Schnitt stabile Effekte (Leichtmann und Nitsch 2020a). Dies bedeutet nicht, dass andere Faktoren unwichtig sind, es zeigt lediglich, dass sich die Wohlfühlidistanz zu einem bestimmten Roboter aus einem komplexen

Zusammenspiel unterschiedlicher Faktoren ergibt. Während ein Faktor in manchen Situationen einen starken Einfluss haben kann, kann er in einer anderen Situation von untergeordneter Wichtigkeit sein.

9.2.5 Ängstlichkeit gegenüber neuen Technologien und Robotern

Ängste gehören zu den unangenehmsten Formen menschlichen Erlebens. Wenn Personen Angst vor Robotern haben, werden sie versuchen, Interaktionen mit Robotern so gut wie möglich zu vermeiden. Ein wesentliches Ziel der Gestaltung sozialer Roboter und ihrer Einsatzgebiete muss es daher sein, dass diese auf ihre menschlichen Nutzer*innen keinesfalls angsteinflößend wirken.

In der Psychologie wird zwischen Angst als Zustand (*state anxiety*) und Ängstlichkeit als Merkmal einer Person (*trait anxiety*) unterschieden (vgl. Spielberger 1983). Obwohl manche Menschen generell, also über viele Bereiche hinweg, ängstlicher sind als andere, wird Ängstlichkeit zumindest teilweise auch als domänenspezifische Eigenschaft betrachtet. Das bedeutet, dass ängstliche Menschen auf unterschiedliche Formen der wahrgenommenen Bedrohung unterschiedlich stark reagieren. Während die eine Person also beispielsweise Angst vor Prüfungen verspürt, hat die andere Person mehr Angst davor, Fremde anzusprechen und die dritte Person wiederum sorgt sich eher vor einer zahnärztlichen Behandlung als vor neuen Sozialkontakten. Mit der wachsenden Prävalenz von Robotern im real erlebten Alltag oder in der medialen Berichterstattung kann sich Ängstlichkeit zunehmend auch gegen Roboter richten.

In Verbindung mit dem Begriff der Roboter-Ängstlichkeit (engl. *robot anxiety*) wird wissenschaftlich untersucht, unter welchen Bedingungen Roboter Gefühle der Angst oder Nervosität hervorrufen und wie sich diese wiederum zum Beispiel auf Einstellungen gegenüber Robotern oder Nutzungsabsichten auswirken (vgl. Nomura und Kanda 2003). In einem Überblicksartikel zum themenspezifischen Stand der Forschung berichten Naneva und Kolleg*innen (2020) von 20 Studien, die sich speziell mit Ängstlichkeit gegenüber sozialen Robotern auseinandergesetzt haben. Zusammenfassend kommen sie zum Schluss, dass bisherige Studienteilnehmer*innen im Mittel angaben, nur wenig Angst vor sozialen Robotern zu haben. 45 Prozent der durchgeführten Studien resultierten sogar in einer neutralen Bewertung (Naneva et al. 2020). Dem gegenüber stehen Einzelstudien wie jene von Liang und Lee (2017), die auf Basis einer repräsentativen Umfrage in den USA nahelegen, dass rund ein Viertel der Bevölkerung erhöhte Ängstlichkeitswerte gegenüber der in diesem Fall breiter definierten Domäne von Robotern und Künstlicher Intelligenz hätte.

Während es im Bereich spezifischer Roboter-Ängstlichkeit zur Klärung zentraler Fragen sicher noch weiterer empirischer Forschung bedarf, liegen im thematisch verwandten Feld der Computer-Ängstlichkeit bereits zahlreiche Befunde aus mehreren Jahrzehnten vor, die wertvolle Anknüpfungspunkte für die Roboterpsychologie liefern können. In einer Zusammenschau von insgesamt 276 wissenschaftlichen Artikeln zum Phänomen der Computer-Ängstlichkeit weist Powell (2013) darauf hin, dass vor allem Faktoren, die mit Erfahrung, Wissen und persönlicher Kompetenzwahrnehmung im Umgang mit Computern zu tun haben, mit besonders geringen Ausprägungen von Computer-Ängstlichkeit assoziiert waren. Im Vergleich etwa zu Alters- oder Geschlechterunterschieden erwiesen sich diese auch als deutlich bedeutsamer. Übertragen auf die Soziale Robotik untermauert eine solche Forschungsbilanz die Notwendigkeit längsschnittlicher Studiendesigns, in denen neben initialen Bewertungen robotischer Systeme auch Gewöhnungs- und Trainingseffekte über längere Zeiträume hinweg betrachtet werden. Im Sinne eines Abbaus von Ängsten untermauern sie daneben die Relevanz praxisnaher Angebote zum spielerischen Erfahrungserwerb im Umgang mit sozialen Robotern sowie einer breit gestreuten Wissensvermittlung und eventuell auch Entmystifizierung dessen, was soziale Roboter sind, wo ihre Chancen und Grenzen liegen und wie die Interaktion mit ihnen funktioniert.

9.3 Methoden der psychologischen Forschung

Wissenschaftliche Erkenntnisse sind untrennbar mit den Methoden zu deren Erlangung verbunden. Um neue Ergebnisse aus der psychologischen Forschung zur Mensch-Roboter-Interaktion richtig einordnen zu können, ist es wichtig, auch die den Ergebnissen zugrundeliegenden Methoden zu kennen und zu verstehen. Psychologische Forschungsmethoden stellen in ihrer Komplexität ein

eigenes Forschungsgebiet dar, das zunehmend an Bedeutung gewinnt. An dieser Stelle soll nichtsdestotrotz der Versuch eines kleinen Überblicks gemacht werden, auf dessen Basis sich interessierte Leser*innen genauer mit entsprechender Fachliteratur auseinandersetzen können (z.B. Döring und Bortz 2016; Schmitz-Atzert und Amelang 2012).

9.3.1 Grundlagen psychologischer Diagnostik

Bei psychologischen Untersuchungen stehen psychologische Konstrukte im Fokus. Damit gemeint sind Merkmale wie Persönlichkeit oder Intelligenz, die nicht unmittelbar beobachtbar, sondern latent sind. Beispiele für Konstrukte, die mit Bezug zur Sozialen Robotik häufig untersucht werden, sind Vertrauen in Roboter (► Abschn. 9.2.3) oder die Zuschreibung von Menschenähnlichkeit (► Abschn. 9.2.1). Konstrukte sind real, aber nicht in dem Sinne, dass sie beispielsweise direkt zählbar sind, sondern als Zusammenfassung von Eigenschaften, die miteinander in Beziehung stehen und durch Zuordnungsregeln (Korrespondenzregeln) beschrieben werden. Durch Konstrukte lassen sich beobachtbare Verhaltensweisen von Individuen erklären und vorhersagen. Dafür müssen sie sehr sorgfältig definiert und messbar gemacht werden. Bei dieser Messbarmachung psychologischer Konstrukte spricht man von Operationalisierung. Hierfür gibt es verschiedene Möglichkeiten.

Psychologische Tests werden beispielsweise häufig als Messmethode zur Erfassung eines psychologischen Merkmals verwendet. Das Ziel ist es dabei, eine quantitative Aussage (eine Ausprägung) über das Merkmal zu treffen, wobei angenommen wird, dass die Variation der Testwerte überwiegend auf die Variation des wahren zu messenden Merkmals zurückzuführen ist (für eine ausführlichere Beschreibung siehe Lehrbücher zur psychologischen Diagnostik, z.B. Schmitz-Atzert und Amelang 2012). Natürlich ist nicht jeder Versuch der Messbarmachung gleich gut.

Psychologische Tests müssen grundsätzlich bestimmten Qualitätskriterien genügen. Diese sind zumeist Objektivität (Wie unabhängig ist ein Testergebnis von dessen Beobachter*in?), Reliabilität (Wie zuverlässig misst ein Test ein Merkmal?) und Validität (Inwiefern misst ein Test das, was er überhaupt messen soll?). Aus diesem Grund durchlaufen psychologische Tests in der Regel aufwändige Konstruktions- und Prüfprozesse zur Bestimmung ihrer Güte.

Ein Beispiel für psychologische Tests wären Selbstberichtsfragebögen, die zur Messung eines latenten Merkmals (z.B. Wie viel Vertrauen bringt eine Person einem Roboter gegenüber?) aus verschiedenen Testfragen oder Testaufgaben (Items) bestehen und von Studienteilnehmer*innen selbst ausgefüllt werden. Solche Fragebögen inkludieren oftmals viele verschiedene Items. Möchte man beispielsweise die Einstellung von Personen gegenüber Robotern messen, so kann die Beurteilung des Statements „Roboter finde ich gut“ (z.B. mit fünf Abstufungen von „trifft nicht zu“ bis „trifft zu“) natürlich schon einen ersten groben Anhaltspunkt geben. Gerade Konstrukte wie Einstellungen sind aber sehr komplex. So könnten Personen es gut finden, dass Maschinen in der Regel sehr exakt arbeiten, sie könnten es aber gleichzeitig als eher negativ empfinden, dass durch neue Maschinen auch Arbeitsplätze gefährdet sein könnten. Für ein tiefes Verständnis sollte ein Konstrukt daher möglichst umfänglich erfasst werden, weil es sonst schnell zu falschen Schlüssen kommen kann. Im Fragebogen von Leichtmann et al. (2021) zur Messung der Einstellung gegenüber kooperativen industriellen Robotern wird neben einer affektiven Evaluation des Roboters (z.B. als gut oder schlecht) beispielsweise auch erfasst, was Personen über mögliche arbeitsbezogene und auch soziale Folgen denken.

9.3.2 Untersuchungsdesigns in der Psychologie

Die Psychologie – und somit auch die Roboterpsychologie – ist eine empirische Wissenschaft. Das heißt, die wissenschaftliche Erkenntnis lässt sich an der Erfahrung überprüfen. Abgeleitet von einer Theorie wird dabei eine Annahme über menschliches Denken, Fühlen und Verhalten durch Beobachtungen, zum Beispiel in Form von Experimenten zur Mensch-Roboter-Interaktion, überprüft (Deduktion). Über solche Beobachtungen können wiederum mithilfe von Abstraktion und Verallgemeinerung auch Rückschlüsse auf die Theorie gezogen werden (Induktion).

In diesem Prozess hat die wissenschaftliche Hypothese einen zentralen Stellenwert. Eine Hypothese ist eine Aussage, die 1) empirisch überprüfbar ist, 2) einen Anspruch der Allgemeingültigkeit hat (über einen Einzelfall hinausgehend), 3) oftmals in Form eines Konditionalsatzes („Wenn-dann“ oder „Je-

desto“) formuliert ist und 4) mithilfe von Daten falsifizierbar ist. Hypothesen beziehen sich auf die Beziehung zwischen verschiedenen Variablen. Variablen sind Mengen von Ausprägungen eines psychologischen Merkmals. Hier wird oft zwischen unabhängigen und abhängigen Variablen unterschieden, wobei versucht wird, die abhängigen Variablen durch die unabhängigen Variablen zu erklären. Ein Beispiel für eine solche Hypothese wäre der Satz „Je mehr positive Erfahrung eine Person mit einem Roboter gesammelt hat, desto mehr Vertrauen bringt sie ihm entgegen“, der einen Zusammenhang zwischen der unabhängigen Variable „Erfahrung“ und der abhängigen Variable „Vertrauen“ annimmt.

Zur Überprüfung solcher Hypothesen wird in der Psychologie eine möglichst repräsentative Zufallsstichprobe aus einer Population oder Grundgesamtheit gezogen. Anhand der Messwerte in dieser Stichprobe sollen dann Rückschlüsse auf die Population gezogen werden. Die Stichprobenkennwerte – und demnach die darauf basierenden Zusammenhänge zwischen Variablen – entsprechen aber nicht den Populationsparametern, sondern sind lediglich Schätzwerte. Diese Schätzer können aber zum Beispiel je nach Güte und Größe der Stichprobe auch deutlich von den wahren Populationskennwerten abweichen. Bei Hypothesen in der Psychologie handelt es sich daher um Wahrscheinlichkeitsaussagen. Sie sind nicht deterministisch, sondern probabilistisch.

Je nach Art der Hypothese können verschiedene Forschungsdesigns für ihre Überprüfung herangezogen werden. Das Experiment stellt das wohl häufigste in der Mensch-Roboter-Interaktion verwendete Untersuchungsdesigns dar (Eyssel 2017). In einem Experiment werden Hypothesen überprüft, indem bestimmte Situationen künstlich in einer kontrollierten Umgebung (zum Beispiel in einer Laborumgebung) herbeigeführt werden. Dabei wird eine unabhängige Variable manipuliert, d.h. in zwei oder mehr Varianten dargeboten. Gemessen wird dann der Effekt dieser verschiedenen Varianten (Bedingungen) auf eine abhängige Variable. Ein Beispiel hierfür wäre ein Vergleich der Vertrauenswerte zwischen einer Experimentalgruppe A, die vor der Vertrauensmessung ein neuartiges Roboter-Kompetenz-Training durchläuft, und einer Kontrollgruppe B, die das herkömmliche Training durchläuft. Mögliche Störvariablen – also Variablen, die auch Varianz in der abhängigen Variablen hervorrufen, die sich jedoch nicht auf die unabhängige Variable zurückführen lassen – werden in Experimenten kontrolliert. Dies soll eine möglichst genaue und wiederholbare Messung ermöglichen. Ein weiteres wichtiges Kriterium eines psychologischen Experiments ist die Zufälligkeit der Aufteilung von Versuchspersonen auf die verschiedenen Versuchsbedingungen. Man spricht hier von Randomisierung.

Eine Randomisierung ist jedoch nicht immer möglich. Möchte man beispielsweise untersuchen, ob ein Interface in Sachen Usability bei Jugendlichen und älteren Menschen gleich gut abschneidet, so kann das Alter nicht zufällig auf Versuchsbedingungen aufgeteilt werden, da es sich hier um eine natürliche Variable handelt. Man spricht dann also nicht mehr von einem Experiment, sondern von einem Quasi-Experiment. Es wird also immer noch die Manipulation kontrolliert, ähnlich wie in einem Experiment, es hat aber keine Randomisierung stattgefunden.

Solche experimentellen Ansätze haben natürlich auch Nachteile. Gerade Laborexperimente werden oft wegen ihrer Künstlichkeit kritisiert. Wird versucht, eine reale Situation im Labor zu simulieren, unterscheidet sich diese häufig doch in vielen Punkten von realen Situationen im Feld (z.B. wenn Versuchspersonen nicht im persönlichen Haushalt auf einen Haushaltsroboter treffen, sondern in einem Büroraum einer Universität). Des Weiteren wissen Versuchspersonen während einer experimentellen Studie, dass sie Teil einer Studie sind und verhalten sich oft anders, als sie dies normalerweise tun würden (siehe z.B. Rosenthal 1966; Orne 1962). Solche Probleme schränken die Validität von experimentellen Forschungsergebnissen ein. Die Erforschung einer Fragestellung sollte daher auf vielen verschiedenen Forschungsansätzen beruhen (Multi-Method-Ansatz). Alternativen beziehungsweise Ergänzungen zu Laborexperimenten sind beispielsweise qualitative Methoden wie Interviews oder Tagebuchstudien, Fragebogenstudien sowie Feldstudien in realen Umgebungen.

9.3.3 Neuere Entwicklungen psychologischer Forschung: Das Erbe einer Krise

2011 war ein schwieriges Jahr für die Psychologie als wissenschaftliche Disziplin, als eine Reihe von Events lostraten, was gemeinhin als „Replizierbarkeitskrise“ bekannt wurde. Es wurde aufgedeckt,

dass durch eine Vielzahl von fragwürdigen Forschungspraktiken (engl. *questionable research practices*, auch bekannt als QRP) praktisch jede auch noch so absurde Hypothese (oft aus Unwissenheit) statistisch bedeutsam gemacht werden kann (Simmons et al. 2011). Infolgedessen wurde in einer Reihe von Studien die Replizierbarkeit psychologischer Experimente untersucht. Das bedeutet, Untersuchungen wurden unter denselben Bedingungen und mit denselben Methoden wie in der ursprünglichen Studie erneut durchgeführt. Es zeigte sich, dass viele Ergebnisse nicht bestätigt werden konnten, vermutlich überschätzt wurden oder sogar gar nicht existieren könnten (Klein et al. 2014; Open Science Collaboration 2015). Diese Replizierbarkeitsprobleme haben verschiedene statistische, theoretische und wissenschaftspraktische Gründe. Im Zuge einer Reformbewegung wurde daraufhin eine Vielzahl an möglichen Lösungen vorgeschlagen. Zu diesen Reformvorschlägen zählen beispielsweise:

- Größere Versuchspersonenzahlen: Einer der Hauptkritikpunkte sind zu kleine Versuchspersonenzahlen in einzelnen Studien, die mehrere Probleme nach sich ziehen können. Beispielsweise kann bei einer zu kleinen Fallzahl auch die Wahrscheinlichkeit klein sein, einen tatsächlich vorhandenen Effekt überhaupt zu finden (Mangel an statistischer Power), oder aber die Stärke von Effekten könnte eher überschätzt werden. Die für eine Studie notwendige Anzahl an Versuchspersonen sollte daher zuvor in einer Power-Analyse statistisch ermittelt werden.
- Größere Transparenz und Open-Science-Praktiken: Durch eine größere Transparenz beim Berichten von Statistiken und Methoden ist es für Leser*innen möglich, besser einzuschätzen, wie Ergebnisse zustande kommen und wie verlässlich diese sind. So ist es zu befürworten, dass beispielsweise Versuchsmaterialien, der Programmcode zur statistischen Auswertung oder die Studiendaten selbst öffentlich zugänglich gemacht werden.
- Präregistrierung: Ein Problem bei Datenanalysen ist die Freiheit an Analysemöglichkeiten, die dazu verleitet, so lange zu testen, bis irgendein Ergebnis signifikant wird. Solch multiples Testen erhöht jedoch die Fehlerrate und somit die Wahrscheinlichkeit eines falsch-positiven Ergebnisses. Registriert man die Studie, also die theoretischen Annahmen und den Auswertungsplan, vor der Studiendurchführung oder der Analyse öffentlich, so schützt man sich selbst vor fragwürdigen Analysepraktiken (QPRs).
- Mehr Replizierung: Die Wiederholung von bereits durchgeführten Studien gibt Aufschluss darüber, wie verlässlich zuvor gefundene Ergebnisse sind. Auch sauber durchgeführte Erhebungen können nämlich zu Zufallsergebnissen führen, da Studien immer nur statistische Wahrscheinlichkeitsaussagen ermöglichen.
- Bessere theoretische Einbettung: Hat sich eine Theorie zuvor in Studien etabliert, kann sie dazu benutzt werden, systematisch abzuleiten, welche Annahmen wahrscheinlicher sind als andere, im Gegensatz zu intuitiven Volksweisheiten oder Laienvermutungen.
- Veröffentlichung nichtsignifikanter Ergebnisse: Oft werden nur die signifikanten Ergebnisse publiziert, nicht jedoch nichtsignifikante Ergebnisse, was zu einer Verzerrung im zugänglichen Korpus an empirischer Literatur führt (*publication bias*). Nur wenn auch nichtsignifikante Ergebnisse publiziert werden, kann sich dem wahren Effekt angenähert werden.

Die Psychologie ist auf einem guten Weg zu verlässlicheren Forschungsergebnissen. Analysen der wissenschaftlichen Befundlage in verschiedenen Forschungsbereichen der Mensch-Roboter-Interaktion (z.B. Leichtmann und Nitsch 2020a; Stower et al. 2021) und einschlägige Replikationsversuche (z.B. Irfan et al. 2018; Leichtmann und Nitsch 2020b; Ullman et al. 2021) deuten jedoch darauf hin, dass die empirische Forschung zu sozialen Robotern in Teilen mit ähnlichen Herausforderungen konfrontiert sein dürfte, mit denen die Psychologie im vergangenen Jahrzehnt umzugehen lernen musste. Eine wachsende Zahl an Forscher*innen kommt daher zum Schluss, dass die oben genannten Reformmaßnahmen auch in künftigen Forschungsprogrammen zu Mensch-Roboter-Interaktion und Sozialer Robotik noch stärkere Berücksichtigung finden müssen (vgl. Belpaeme 2020).

9.4 Brauchen Roboter Psychotherapie?

Der menschenähnliche, emotional empfindsame Roboter auf der Therapiecouch – das ist ein Bild, das wohl so einige spontan mit „Roboterpsychologie“ assoziieren. Ursprünglich stammt der englische

Begriff „Robopsychology“ aus der Feder des Science-Fiction-Autors Isaac Asimov, der die fiktive Robopsychologin Susan Calvin in seinem populären Buch „I, Robot“ tatsächlich als Betreuerin intelligenter humanoider Roboter einführte. In der Realität befasst sich die Roboterpsychologie, wie zu Beginn des Kapitels dargestellt, natürlich nicht mit dem Wohlbefinden von Robotern, sondern mit dem Erleben und Verhalten jener Menschen, die in ihrem Alltags- und Berufsleben zunehmend mit intelligenten Maschinen in der Rolle sozialer und kollaborativer Agenten konfrontiert sind. Daneben gilt es anzumerken, dass Roboter – und seien sie noch so überzeugend menschengleich gestaltet – klarerweise keine lebendigen organischen Wesen mit Ichbewusstsein oder realer emotionaler Erlebensfähigkeit darstellen und ihr Bedarf an psychotherapeutischer Unterstützung schon aus diesem Grund bis auf weiteres auf die Science-Fiction-Sphäre beschränkt bleiben wird. Gleichzeitig bedeutet das aber nicht, dass Maschinen nicht trotzdem häufig eine Art „Psyche“ zugeschrieben würde oder dass Roboter nicht als neue technologische Tools in therapeutischen Kontexten getestet würden, wie in den folgenden abschließenden Abschnitten dargestellt wird.

9.4.1 Zur „Psyche“ des Roboters

Die Psycholog*innen Fritz Heider und Marianne Simmel führten im Jahr 1944 ein interessantes Experiment durch. Im Rahmen einer Lehrveranstaltung präsentierten sie ihren Studierenden eine simple Schwarz-Weiß-Animation, in der sich ein kleines Dreieck, ein größeres Dreieck und ein Kreis in unterschiedlichen Weisen aufeinander zu- und voneinander wegbewegten. Im Anschluss daran sollten die Studierenden nacherzählen, was sie in dem Stummfilm gesehen hatten (Heider und Simmel 1944). Das Spannende: Die meisten beschrieben nicht etwa zweidimensionale Positionsänderungen, sondern erzählten von den Zielen, Gefühlslagen und Beziehungen der geometrischen Figuren, die sie in deren animierten Bewegungen wahrgenommen hatten. Das kleine Dreieck und der Kreis wurden beispielsweise als „Freunde“ beschrieben, das große Dreieck als „wütend“, und so weiter.

Diese frühe Studie und auch diverse neuere Forschungsergebnisse deuten darauf hin, dass es manchmal gar nicht viel braucht, um nichtmenschlichen Entitäten gleichsam automatisiert innere Gedanken, Gefühle und Motivationen zuzuschreiben, selbst wenn uns eigentlich klar ist, dass diese in realitas über nichts dergleichen verfügen (vgl. dazu auch  Abschn. 9.2.3). Ganz besonders leicht fällt uns die Zuschreibung von Psyche aber wohl bei Akteuren, die speziell darauf programmiert wurden, „Persönlichkeit“ oder „Gefühle“ nach außen darzustellen. Ein Beispiel hierfür sind sogenannte emotionale Roboter, die einerseits affektive Zustände ihrer menschlichen Interaktionspartner*innen erkennen sollen und andererseits durch simplen mimischen, gestischen oder sprachlichen Ausdruck vordefinierte „Emotionen“ auch selbst zeigen. Wenn der Roboter „Pepper“ (SoftBank Robotics) etwa gelobt wird, kommuniziert er, dass er „glücklich“ ist, wenn das Licht ausgeht, signalisiert er „Angst“ (SoftBank Robotics 2015). Solche (häufig noch sehr rudimentär ausgeprägten) Gefühls- oder Empathiesimulationen emotionaler Roboter sollen dem Zweck einer intuitiveren, flüssigeren und sozial akzeptableren Mensch-Maschine-Kommunikation dienen.

Daneben wird eine künstliche Modellierung emotionaler Reaktionen in der Robotik und Künstlichen Intelligenz aber zunehmend auch deswegen als Ziel formuliert, weil zahlreiche psychologische Forschungsarbeiten auf das untrennbare Zusammenspiel von emotionalen und kognitiven Prozessen für schnelle und gewinnbringende Entscheidungsprozesse beim Menschen hindeuten.

Wissenschaftler*innen beschäftigen sich daher mit der Frage, was sich Computerprogramme von menschlichen Emotionen „anschauen“ können, um möglichst effizient und nützlich entscheiden und handeln zu können. Ein Beispiel hierfür wäre die Programmierung einer funktionalen „Angst“-Reaktion, um in einer für den Roboter oder seine*r menschlichen Interaktionspartner*in gefährlichen Situation schnell und schützend agieren zu können. Gleichzeitig bleibt anzuerkennen, dass Roboter trotz von Menschen einprogrammierter oder durch simulierte Erfahrungen antrainierter Emotionssysteme keine „echten Gefühle“ haben, weil die Komponenten des Ichbewusstseins und emotionalen Erlebens letztendlich fehlen (Picard 2003).

9.4.2 Soziale Roboter als Werkzeuge der Psychotherapie

Auch wenn Roboter nicht wahrhaftig empfindsam sind und empathische Interaktionsformen daher nur simulieren können, wird seit einigen Jahren über das Potenzial sozialer Roboter für psychologisch-

diagnostische und psychotherapeutische Zielsetzungen diskutiert. Bisher erprobte Einsatzbereiche betreffen vor allem Stressreduktion, Motivationsaufbau oder die Begleitung verhaltenstherapeutischer Ansätze, in denen strukturierte Übungen oft besonders gut technologisch unterstützt werden können. In empirischen Studien zur Förderung psychischer Gesundheit durch tierähnliche Roboter wie der interaktiven Robbe Paro (Shibata 2012), die auf Streicheln und Geräusche reagiert, oder dem Roboterbären eBear (Zhang et al. 2014) zeigten sich durch gezielte Interaktion mit Altenheimbewohner*innen und teils auch mit Demenzpatient*innen emotional unterstützende Effekte, die sich durch Verringerungen von Angst, Aggression und Depressivität, eine allgemeine Stimmungsverbesserung oder Aktivitätssteigerung ausdrückten (vgl. Eichenberg 2020; Scoglio et al. 2019).

Daneben liegen auch Hinweise auf eine positive Beeinflussung kognitiver und sozialer Prozesse durch Roboter vor. Im Bereich der kognitiven Unterstützung konnte etwa gezeigt werden, dass Interaktionen mit dem kleinen menschenähnlichen Roboter Nao zu einer verbesserten Aufgabenbewältigung bei Demenzkranken (Costescu et al. 2015) und zu einer höheren Motivation, Aufgaben zu lösen, bei Studierenden führten (Galvão Gomes da Silva et al. 2018). Zur Förderung sozialer Prozesse wurden Humanoide wie Nao oder Zeno bisher insbesondere im Rahmen von Verhaltens- und Fertigkeitstrainings für autistische Personen ausprobiert, beispielsweise um autistische Kinder zur sozialen Interaktion zu animieren oder um spielerische Lernerfahrungen mit mimischen Gefühlsausdrücken zu ermöglichen und Imitationsfähigkeiten zu verbessern. Verhaltensformen, die in Interaktionsspielen mit Robotern neu erlernt wurden, können in darauffolgenden Sitzungen mit menschlichen Therapeut*innen dann weiter verbessert und verfeinert werden (vgl. Eichenberg 2020).

Ein besonders kontrovers diskutiertes Anwendungsfeld robotergestützter Therapieansätze stellt daneben die Sexualtherapie dar. Während einige Autor*innen positive psychosoziale Auswirkungen menschenähnlicher Sexroboter etwa für Personen mit sexuellen Störungen, verstärkter sozialer Ängstlichkeit oder mit physischen oder kognitiven Beeinträchtigungen annehmen (für eine kurze Übersicht siehe Döring 2020), streichen andere Forscher*innen potenzielle Risiken in Bezug auf die Privatsphäre und Sicherheit der Nutzer*innen, die emotionale Manipulation verletzlicher Personen, die Objektivierung des menschlichen Körpers, die Perpetuierung stereotyper geschlechtsspezifischer Rollenmuster oder die Förderung sexueller Gewalt heraus (vgl. Appel et al. 2019).

Insgesamt liegen für den Einsatz unterschiedlicher sozialer Roboter in psychotherapeutischen Kontexten zwar erste empirische Hinweise auf positive Wirkungen vor, gleichzeitig ist das Feld aber noch stark im Aufbau begriffen. Nicht zuletzt in Anbetracht der teils noch unklaren ethischen Implikationen können Therapieroboter in jedem Fall nur als ergänzende Werkzeuge für die Arbeit menschlicher Therapeut*innen betrachtet werden. Erst im Rahmen eines Therapieprogramms mit ausgebildeten (Psycho-)Therapeut*innen kann abgeschätzt werden, ob der Einsatz von Robotern für bestimmte Personen geeignet ist. Um Risiken aufgeklärt beurteilen und Potenziale für Patient*innen bestmöglich bewerten zu können, wäre eine verstärkte Wissens- und Kompetenzvermittlung zu sozialen Robotern in Richtung junger Therapeut*innen wünschenswert. In einer Umfrage mit europäischen Psychologiestudierenden gab ein Großteil der Befragten an, nicht über die notwendigen Fähigkeiten zu verfügen, um einen Roboter zukünftig sinnvoll für eine Tätigkeit als Therapeut*in einsetzen zu können (Conti et al. 2019). Die Stärkung eines fächerübergreifenden Austauschs zwischen den Technikwissenschaften, den Human- und Sozialwissenschaften, der Ethik und anderen Disziplinen wäre hier – wie auch in vielen weiteren Teildomänen der Sozialen Robotik – durchwegs begrüßenswert.

9.5 Literatur

- Aiello JR (1987) Human spatial behavior. In: Stokols D., Altman I (Hrsg) *Handbook of Environmental Psychology*. Wiley, New York
- Appel M, Marker C, Mara M (2019) Otakuism and the appeal of sex robots. *Frontiers in psychology*, 10:569
- Appel M, Izydorczyk D, Weber S, Mara M, Lischetzke T (2020) The uncanny of mind in a machine: Humanoid robots as tools, agents, and experiencers. *Computers in Human Behavior*, 102:274-28
- Appel M, Weber S, Krause S, Mara M (2016, March) On the eeriness of service robots with emotional capabilities. In: 2016 11th ACM/IEEE International Conference on Human-Robot Interaction (HRI), S 411–412
- Bartneck C, Kanda T, Ishiguro H, Hagita N (2009) My robotic doppelgänger—A critical look at the uncanny valley. The 18th IEEE International Symposium on Robot and Human Interactive Communication, Toyama, Japan, S 269–276
- Belpaeme T (2020) Advice to New Human-Robot Interaction Researchers. In: Jost C, Le Pévédic B, Belpaeme T, Bethel C, Chrysostomou D, Crook N, Grandgeorge M, Mirnig N (Hrsg) *Human-Robot Interaction. Evaluation methods and their standardization*. Springer International Publishing, Cham, S 355–369
- Burleigh TJ, Schoenherr JR, Lacroix GL (2013) Does the uncanny valley exist? An empirical test of the relationship between eeriness and the human likeness of digitally created faces. *Computers in Human Behavior*, 29:759–771
- Conti D, Cattani A, Di Nuovo S, Di Nuovo A (2019) Are future psychologists willing to accept and use a humanoid robot in their practice? Italian and English students' perspective. *Frontiers in psychology*, 10:2138
- Costescu CA, Vanderborcht B, David DO (2015) Reversal learning task in children with autism spectrum disorder: A robot-based approach. *Journal of Autism and Developmental Disorders*, 45:3715–3725
- De Visser EJ, Peeters MM, Jung MF, Kohn S, Shaw TH, Pak R, Neerincx MA (2020) Towards a theory of longitudinal trust calibration in human–robot teams. *International journal of social robotics*, 12:459–478
- Döring N (2020) Sexpuppen und Sexroboter aus psychologischer und therapeutischer Perspektive; Pathogene und salutogene Nutzungsmuster. In: Bendel O (Hrsg) *Maschinenliebe. Liebespuppen und Sexroboter aus technischer, psychologischer und philosophischer Perspektive*. Springer Fachmedien Wiesbaden, Wiesbaden, S 283–301
- Döring N, Bortz J (2016) *Forschungsmethoden und Evaluation in den Sozial- und Humanwissenschaften*. 5. Aufl. Springer, Berlin und Heidelberg
- Echterhoff G, Bohner G, Siebler F (2006) „Social Robotics“ und Mensch-Maschine-Interaktion. *Zeitschrift für Sozialpsychologie*, 37:219-231
- Eichenberg C (2020) Robotik in der Psychotherapie: Anwendungsfelder–Effektivität–Praxisbeispiele. In *Bessere Menschen? Technische und ethische Fragen in der transhumanistischen Zukunft* (pp. 97-125). Springer, Berlin, Heidelberg
- Epley N, Waytz A, Cacioppo JT (2007) On seeing human: a three-factor theory of anthropomorphism. *Psychological review*, 114(4), 864
- Esterwood C, Essenmacher K, Yang H, Zeng F, Robert L (2021) A Meta-Analysis of Human Personality and Robot Acceptance in Human-Robot Interaction. In *Conference on Human Factors in Computing Systems (CHI'21)*, May 8–13, 2021, Yokohama, Japan. ACM, New York (New York)
- Eyssel F, Reich N (2013) Loneliness makes the heart grow fonder (of robots)—on the effects of loneliness on psychological anthropomorphism. In: *8th ACM/IEEE international conference on human–robot interaction (HRI)*, S 121–122
- Eyssel F (2017) An experimental psychological perspective on social robotics. *Robotics and Autonomous Systems*, 87:363–371
- Galvão Gomes da Silva J, Kavanagh DJ, Belpaeme T, Taylor L, Beeson K, Andrade J (2018) Experiences of a motivational interview delivered by a robot: Qualitative study. *Journal of Medical Internet Research*, 20(5):e116
- Gerrig RJ (2016) *Psychologie*. 20., aktualisierte und erweiterte Auflage. Pearson Studium, Hallbergmoos
- Gifford R (2013) Personal Space and Territoriality. In: Gifford R (Hrsg) *Environmental psychology: Principles and Practice*, Optimal Books, Victoria, S 123–164
- Heider F, Simmel M (1944) An experimental study of apparent behavior. *American Journal of Psychology*, 57:243–259
- Hall ET (1966) *The hidden dimension*. Doubleday, Garden City(New York)
- John OP, Srivastava S (1999) The Big-Five trait taxonomy: History, measurement, and theoretical perspectives (Vol. 2). University of California, Berkeley, 102–138
- Kätsyri J, Förger K, Mäkäräinen M, Takala T (2015) A review of empirical evidence on different uncanny valley hypotheses: support for perceptual mismatch as one road to the valley of eeriness. *Frontiers in psychology*, 6:390
- Koerber M (2018) Theoretical considerations and development of a questionnaire to measure trust in automation. In: *Congress*

- Leichtmann B, Hartung J, Wilhelm O, Nitsch V (2021) Measuring attitudes toward cooperative robots in industrial work settings: Development of a new questionnaire and exploration of attitude structure. Manuscript submitted for publication
- Leichtmann B, Nitsch V (2020a) How much distance do humans keep toward robots? Literature review, meta-analysis, and theoretical considerations on personal space in human-robot interaction. *Journal of Environmental Psychology*, 68:101386
- Leichtmann B, Nitsch V (2020b) Is the Social Desirability Effect in Human–Robot Interaction overestimated? A Conceptual Replication Study Indicates Less Robust Effects. *International Journal of Social Robotics*:1–19
- Lischetzke T, Izydorczyk D, Hüller C, Appel M (2017) The topography of the uncanny valley and individuals' need for structure: a nonlinear mixed effects analysis. *Journal of Research in Personality*, 68:96–113
- MacDorman KF, Entezari SO (2015) Individual differences predict sensitivity to the uncanny valley. *Interaction Studies*, 16:141–172
- Mara M, Meyer K, Heiml M, Pichler H, Haring R, Krenn B, Gross S, Reiterer B, Layer-Wagner, T (2021) CoBot Studio VR: A Virtual Reality Game Environment for Transdisciplinary Research on Interpretability and Trust in Human-Robot Collaboration. In: *Int. Workshop on Virtual, Augmented, and Mixed Reality for HRI (VAM-HRI), HRI'21, Boulder*
- Mara M, Appel M (2015a) Effects of lateral head tilt on user perceptions of humanoid and android robots. *Computers in Human Behavior*, 44:326–334
- Mara M, Appel M (2015b) Science fiction reduces the eeriness of android robots: A field experiment. *Computers in Human Behavior*, 48:156–162
- Mathur MB, Reichling, DB (2016) Navigating a social world with robot partners: A quantitative cartography of the Uncanny Valley. *Cognition*, 146:22-32
- McAllister DJ (1995) Affect-and cognition-based trust as foundations for interpersonal cooperation in organizations. *Academy of management journal*, 38:24–59
- Mori M (1970) Bukimi no tani [The uncanny valley]. *Energy*, 7:33–35
- Naneva S, Sarda Gou M, Webb TL, Prescott TJ (2020) A systematic review of attitudes, anxiety, acceptance, and trust towards social robots. *International Journal of Social Robotics*:1–23
- Nass C, Moon Y, Carney P (1999) Are people polite to computers? Responses to computer-based interviewing systems. *Journal of applied social psychology*, 29:1093-1109
- Nomura T, Kanda T (2003, November) On proposing the concept of robot anxiety and considering measurement of it. In: *The 12th IEEE International Workshop on Robot and Human Interactive Communication, 2003. Proceedings. ROMAN 2003*, S 373–378
- Orne MT (1962) On the social psychology of the psychological experiment: With particular reference to demand characteristics and their implications. *American Psychologist*, 17:776–783
- Picard RW (2003) What does it mean for a computer to “have” emotions. *Emotions in humans and artifacts*:213–235
- Powell AL (2013) Computer anxiety: Comparison of research from the 1990s and 2000s. *Computers in Human Behavior*, 29:2337–2381
- Reeves B, Nass C (1996) *The media equation: How people treat computers, television, and new media like real people.* Cambridge University Press, Cambridge (UK)
- Rempel JK, Holmes JG, Zanna MP (1985) Trust in close relationships. *Journal of personality and social psychology*, 49: 95
- Robert L, Alahmad R, Esterwood C, Kim S, You S, Zhang Q (2020) A Review of Personality in Human–Robot Interactions. Available at SSRN 3528496
- Roese NJ, Amir E (2009) Human-Android Interaction in the Near and Distant Future. *Perspectives on Psychological Science*, 4:429–434
- Rosenthal R (1966) *Experimenter effects in behavioral research: Appleton-Century-Crofts*
- Salem M, Lakatos G, Amirabdollahian F, Dautenhahn K (2015) Would You Trust a (Faulty) Robot?: Effects of Error, Task Type and Personality on Human-Robot Cooperation and Trust, In *Proceedings of the ACM/IEEE International Conference on Human-Robot Interaction (HRI'15)*, Portland, OR, S 141–148
- Schmidt-Atzert L, Amelang M (2012) *Psychologische Diagnostik. 5. Aufl.* Springer, Berlin und Heidelberg
- Shamsuddin S, Yussof H, Ismail L, Hanapiah FA, Mohamed S, Piah HA, Zahari NI (2012, March) Initial response of autistic children in human-robot interaction therapy with humanoid robot NAO. In: *2012 IEEE 8th International Colloquium on Signal Processing and its Applications*, S 188–193

-
- Shibata T (2012) Therapeutic seal robot as biofeedback medical device: Qualitative and quantitative evaluations of robot therapy in dementia care. *Proceedings of the IEEE*, 100, S 2527–2538
- Smith ER, Šabanović S, Fraune MR (2021) Human-Robot Interaction Through the Lens of Social Psychological Theories of Intergroup Behavior. *Technology, Mind, and Behavior*, 1(2). <https://tmb.apaopen.org/pub/v1ke08m8>. Zugegriffen: 14. März 2021
- Scoglio AA, Reilly ED, Gorman JA, Drebing CE (2019). Use of social robots in mental health and well-being research: Systematic review. *Journal of Medical Internet Research*, 21:e13322
- SoftBank Robotics (2015) SoftBank to Launch Sales of 'Pepper' — the World's First Personal Robot That Reads Emotions. https://www.softbank.jp/en/corp/group/sbm/news/press/2015/20150618_01/. Zugegriffen: 29. März 2021
- Spielberger CD (1983) State-Trait anxiety inventory. *The Corsini encyclopedia of psychology*, 1–1
- Stein JP, Ohler P (2017) Venturing into the uncanny valley of mind—The influence of mind attribution on the acceptance of human-like characters in a virtual reality setting. *Cognition*, 160:43-50
- Stower R, Calvo-Barajas N, Castellano G, Kappas A (2021) A meta-analysis on children's trust in social robots. *International Journal of Social Robotics*:1-23
- Ullman D, Aladia S, Malle, BF (2021) Challenges and Opportunities for Replication Science in HRI: A Case Study in Human-Robot Trust. In: *HRI '21: Proc. of the 2021 ACM/IEEE International Conference on Human-Robot Interaction*, S 110–118
- Waytz A, Cacioppo J, Epley N (2010) Who sees human? The stability and importance of individual differences in anthropomorphism. *Perspectives on Psychological Science*, 5:219–232
- Weir K (2018) The dawn of social robots. Roboticians must dip heavily into psychological science. *Monitor on Psychology*, 49:50
- Wolfert P, Deschuyteneer J, Oettringer D, Robinson N, Belpaeme T (2020, March) Security risks of social robots used to persuade and manipulate: A proof of concept study. In: *Companion of the 2020 ACM/IEEE International Conference on Human-Robot Interaction*, S 523–525
- Zhang X, Mollahosseini A, Kargar B. AH, Boucher E, Voyles RM, Nielsen R, Mahoor MH (2014) eBear: An expressive Bear-Like robot. *23rd IEEE International Symposium on Robot and Human Interactive Communication*, Edinburgh, UK, 2014, S 969–974