

Tobias Kopp

VERTRAUEN IN ROBOTER UND DESSEN BEEINFLUSSBARKEIT DURCH SPRACHLICHES FRAMING

*Eine empirische Untersuchung der Interaktion
mit Cobots am Arbeitsplatz*



Tobias Kopp

Vertrauen in Roboter und dessen Beeinflussbarkeit durch sprachliches Framing

Eine empirische Untersuchung der Interaktion mit Cobots am Arbeitsplatz

Vertrauen in Roboter und dessen Beeinflussbarkeit durch sprachliches Framing

Eine empirische Untersuchung der Interaktion
mit Cobots am Arbeitsplatz

von
Tobias Kopp

Karlsruher Institut für Technologie
Institut für Technikfolgenabschätzung und Systemanalyse

Vertrauen in Roboter und dessen Beeinflussbarkeit durch
sprachliches Framing: Eine empirische Untersuchung der
Interaktion mit Cobots am Arbeitsplatz

Zur Erlangung des akademischen Grades eines Doktors der Philosophie
(Dr. phil.) von der KIT-Fakultät für Geistes- und Sozialwissenschaften des
Karlsruher Instituts für Technologie (KIT) genehmigte Dissertation

von Tobias Kopp

Tag der mündlichen Prüfung: 14. Februar 2022

Erster Gutachter: Prof. Dr. Armin Grunwald

Zweiter Gutachter: Prof. Dr. Steffen Kinkel

Bild Umschlagvorderseite: Blue Planet Studio/Shutterstock.com

Bild Umschlagrückseite: Tobias Schwerdt

Impressum



Karlsruher Institut für Technologie (KIT)

KIT Scientific Publishing

Straße am Forum 2

D-76131 Karlsruhe

KIT Scientific Publishing is a registered trademark
of Karlsruhe Institute of Technology.

Reprint using the book cover is not allowed.

www.ksp.kit.edu



*This document – excluding parts marked otherwise, the cover, pictures and graphs –
is licensed under a Creative Commons Attribution 4.0 International License
(CC BY 4.0): <https://creativecommons.org/licenses/by/4.0/deed.en>*



*The cover page is licensed under a Creative Commons
Attribution-NonCommercial 4.0 International License (CC BY-NC 4.0):
<https://creativecommons.org/licenses/by-nd/4.0/deed.en>*

Print on Demand 2022 – Gedruckt auf FSC-zertifiziertem Papier

ISBN 978-3-7315-1206-6

DOI 10.5445/KSP/1000146827

Kurzfassung

Moderne Roboter fungieren zunehmend als vielfältige künstliche Begleiter:innen der Menschen und kreieren neue Qualitäten der Mensch-Roboter-Beziehung. Auch im Bereich der industriellen Arbeit als klassischer Einsatzbereich von Robotern stellen sog. kollaborationsfähige Roboter (Cobots) bestehende Relationen auf den Prüfstein. Diese Cobots verfügen über implementierte Sicherheitsmechanismen und erlauben damit eine Mensch-Roboter-Interaktion (MRI) inklusive eines physischen Kontakts ohne trennende Schutzvorrichtungen. Dadurch verschwimmen die Grenzen zwischen Robotern als klassische Werkzeuge und als soziale Interaktionspartner:innen. Das erzeugt häufig ambivalente Einstellungen auf Seiten der Mitarbeitenden, die wiederum durch das am Arbeitsplatz typische Spannungsfeld zwischen der Ergänzung oder der Ersetzung menschlicher Arbeit durch Technologie zusätzlich befördert werden. Die Möglichkeit einer kollaborierenden Verwendung von Cobots, bei der die Mitarbeitenden im Fertigungsprozess integriert bleiben, kollidiert mit bestehenden mentalen Modellen von Robotern, die häufig einer Rationalisierungslogik folgen. Dadurch ist es gleichermaßen naheliegend, dass sich Mitarbeitende entweder in einer kooperierenden oder in einer konkurrierenden Relation zum Cobot verorten. Häufig verwendete anthropomorphe Narrative, Metaphern und Designmerkmale von Cobots erschweren zusätzlich eine eindeutige mentale Kategorisierung und lassen Cobots entweder als menschenähnliche Kolleg:innen oder als bloße Maschinen erscheinen. Angesichts dieser Mehrdeutigkeiten lässt sich vermuten, dass sprachliches Framing, das auf unterschiedlichen Arten der Informationsvermittlung basiert, imstande ist, die mentalen Modelle von Cobots und das Vertrauen von Mitarbeitenden in Cobots zu beeinflussen. Dieser vermutete Einfluss unter Berücksichtigung der Spezifika des Arbeitsplatzkontexts ist bisher kaum untersucht und stellt daher das Forschungsgebiet dieser Dissertation dar. In Ermangelung übergreifender Theorien zum Mensch-Cobot-Vertrauen am Arbeitsplatz und zur Wirkung des sprachlichen Framings werden Theorieansätze aus unterschiedlichen Disziplinen herangezogen. Dazu zählen bspw. das Modell zum Vertrauen in Automation von J. D. Lee und See (2004), das Drei-Schichten-Modell von Hoff und Bashir (2015), die soziologische Vertrauenskonzeption nach Luhmann (1968/2014), die Drei-Faktoren-Theorie des Anthropomorphismus nach Epley, Waytz und Cacioppo (2007) und das Modell der Frame-Selektion nach Esser (2005).

Die Ergebnisse aus einer quantitativen und einer qualitativen Vorstudie verdeutlichen, dass Unternehmensmitarbeitende sowohl dem initialen Vertrauen im Vorfeld einer MRI als auch dem dynamischen Vertrauen während einer MRI eine hohe Relevanz in Hinblick auf den Erfolg einer betrieblichen Cobot-Einführung zuschreiben. Insbesondere die Angst vor Arbeitsplatzverlust kristallisiert sich als omnipräsente und existenzielle Sorge unter den Mitarbeitenden heraus, die die Vertrauensbildung unterminieren kann. Letztere hängt im

Unternehmenskontext stets auch von der wahrgenommenen Vertrauenswürdigkeit der Personen und Institutionen ab, die die Entwicklung, Programmierung und Einführung eines Cobots durchführen. Ferner offenbart sich einerseits ein hohes Vertrauen in die Sicherheit eines Cobots und andererseits eine große Verunsicherung in Hinblick auf die direkten Auswirkungen auf die eigenen Arbeitsroutinen und auf den Erhalt des eigenen Arbeitsplatzes. Entsprechend zeigen Produktionsmitarbeitende eine hohe Sensitivität für all jene Informationen, die sich als Hinweise auf die Leistungsfähigkeit eines Cobots, dessen Einsatzzweck und die Mensch-Cobot-Relation interpretieren lassen. Während die Mitarbeitenden Anthropomorphismus in diesem Kontext als irrelevant beschreiben, deuten sprachliche Merkmale auf implizite Vermenschlichungsprozesse hin.

Die Ergebnisse einer durchgeführten Experimentalstudie in einem Online-Setting legen erwartungskonform nahe, dass sprachliches Framing die wahrgenommene Menschenähnlichkeit eines Cobots und die wahrgenommene Kooperationsorientierung der Mensch-Cobot-Relation beeinflusst. Diese beiden Faktoren wirken sich wiederum auf das initiale Mensch-Cobot-Vertrauen aus. Dabei moderiert die Mensch-Cobot-Relation die Effekte der Vermenschlichung. Ein hohes Maß an Vermenschlichung wirkt sich demnach nur positiv auf das Vertrauen in den Cobot aus, wenn sich die Mitarbeitenden in einer kooperierenden Relation zu einem Cobot sehen. Das Ausbleiben dieses Effekts im Falle einer konkurrierenden Relation verweist auf die unterschiedlichen psychischen Reaktionen auf die vermutete Ersetzung durch Menschen oder Maschinen. Allerdings scheint die Wirkkraft des Framings von der Stichprobe und/oder dem experimentellen Setting abzuhängen, wie eine zweite Experimentalstudie offenbarte, die als Präsenzexperiment inklusive einer realen MRI stärker affektive Vertrauenskomponenten adressierte und mit Studierenden durchgeführt wurde, die sich in ihrer Einstellung und ihren Vorkenntnissen in Bezug auf Roboter deutlich unterschieden. Im Gegensatz zu Experimentalstudie 1 beeinflusste weder das Framing die Wahrnehmung der Studierenden signifikant, noch zeigte sich ein vertrauenssteigernder Effekt durch die Bereitstellung relevanter Informationen über den Cobot. Allerdings beförderte die reale Interaktion mit dem Cobot das Vertrauen erheblich. Vertrauensveränderungen gingen dabei nicht immer mit gegenläufigen Misstrauensveränderungen einher.

Die wahrgenommene Ambiguität des Cobots scheint eine entscheidende Rolle für die Vertrauensbildungsprozesse und die Wirkkraft des sprachlichen Framings zu spielen. Die mit diesen Mehrdeutigkeiten verbundene kognitive Komplexität aufgrund entstehender mentaler Kategorisierungs- und Selektionsprobleme lässt sich in Anlehnung an Luhmann durch Vertrauen als Festlegung auf eine konkrete Zukunftserwartung auflösen. Die vermittelten Frames dienen dabei als diskriminierender Faktor, um eine konkrete Zukunftserwartung auszuwählen. Die personenabhängig unterschiedliche Wirkkraft lässt sich anhand individueller Faktoren wie der bestehenden Vorkenntnisse, der Motivation zur elaborierten Informationsverarbeitung oder der Ambiguitätsaversion erklären. Diese Erklärungsansätze

bedürfen einer weiteren Vertiefung und empirischen Überprüfung. Unabhängig von Framing-Einflüssen gilt die Vertrauensentwicklung als ein zeitlich in verschiedenen Phasen ablaufender dynamischer und multidimensionaler Prozess, der sich u. a. in leistungs- und relationsorientierte Komponenten unterteilen lässt. Ein zentraler Einflussfaktor auf die Vertrauensbildung stellt der Grad an Vermenschlichung dar. Daraus ergeben sich verschiedenartige Erwartungen, die sich wiederum kontextspezifisch unterschiedlich auf die Vertrauensdimensionen auswirken. Die Effekte des Anthropomorphismus auf das Vertrauen werden zusätzlich durch die Mensch-Cobot-Relation moderiert, die sich u. a. durch das Spannungsfeld zwischen Kooperation und Konkurrenz charakterisieren lässt und deutlich mehr Facetten aufweist, als die oftmals thematisierte Dichotomie zwischen der Repräsentation eines Roboters entweder als Werkzeug oder als Kolleg:in nahelegt. Die Differenz zwischen Kooperations- oder Konkurrenzorientierung kann aufgrund der weit verbreiteten Angst vor Arbeitsplatzverlust als entscheidendes Merkmal der Relation am Arbeitsplatz herangezogen werden, sollte aber in künftiger Forschung um weitere qualitative Merkmale von Relationen ergänzt werden. Entscheidend sind hierbei phänomenologische Fragestellungen nach der Wahrnehmung des Roboters, die ontologische Fragen nach dem Status und den tatsächlichen Fähigkeiten desselben zunehmend verdrängen.

Aus den Erkenntnissen ergeben sich unternehmerische sowie gesellschaftliche Implikationen, die sich partiell auch auf andere Technologien und Anwendungskontexte übertragen lassen. Eine angemessene Vertrauenskalibrierung erscheint dabei als ein aus vielerlei Hinsicht wünschenswertes Ziel, das eine achtsame Sprachverwendung in der Rede über neue komplexe Technologien erfordert. Der multimethodische, anwendungsnahe und interdisziplinäre Forschungsansatz erwies sich als zielführend für die Untersuchung des vielschichtigen Mensch-Roboter-Vertrauens, wenngleich die empirischen Ergebnisse u. a. in Hinblick auf die Operationalisierung der Mensch-Cobot-Relation oder die Betrachtung längerfristiger Vertrauensprozesse limitiert sind. Die mangelnde Anwendungsnähe und Kontextsensitivität vieler Studien erschwert die Synthese bestehender experimenteller Ergebnisse und die Entwicklung ganzheitlicher und praktisch anwendbarer Theorien des Mensch-Cobot-Vertrauens am Arbeitsplatz. Letzteres stellt damit ein übergeordnetes Ziel für die weitere Forschung dar. Die vorliegende Dissertation unterstreicht die Notwendigkeit, bestehende experimentelle Ergebnisse und theoretische Ansätze aus verschiedenen Disziplinen wie etwa der Psychologie, Soziologie und Philosophie zu integrieren, um die zahlreichen Facetten, kontextspezifischen Determinanten und Auswirkungen von Vertrauen am Arbeitsplatz ausreichend zu berücksichtigen.

Abstract

Modern robots are increasingly functioning as multifaceted artificial companions for humans and creating new qualities in the human-robot relationship. Also in the field of industrial labor as a classic area of application for robots, so-called collaborative robots (cobots) are putting existing relationships to the test. Due to implemented safety mechanisms, cobots allow human-robot interaction (HRI) including physical contact without separating physical protection. This blurs the boundaries between robots as classic tools or as social interaction partners, often evoking ambivalent attitudes of employees. This ambivalence is further promoted in workplace settings by the typical area of conflict of either supplementing or replacing human work with technology. The option of using cobots in a collaborative manner, in which employees remain integrated in the production process, collides with existing mental models of robots, which often follow a rationalization logic. There is an equal chance that employees locate themselves either in a cooperating or in a competing relation to the cobot. Frequently used anthropomorphic narratives, metaphors, and design features of cobots further complicate a clear mental categorization and make cobots appear either as human-like colleagues or as inhumane machines. Given these ambiguities, it can be assumed that linguistic framing based on different ways of conveying information is able to influence employees' mental models of cobots and their trust in cobots. This presumed influence, taking into account the specifics of the context of the workplace, has hardly been investigated and therefore represents the research area of this dissertation. Due to the lack of a comprehensive theory on human-cobot trust in the workplace and on the effect of linguistic framing, theoretical approaches from different disciplines are applied. These include, for example, the model of trust in automation by J. D. Lee & See (2004), the three-layered model by Hoff & Bashir (2015), the sociological concept of trust by Luhmann (1968/2014), the three-factor theory of anthropomorphism by Epley, Waytz & Cacioppo (2007) and the model of frame selection by Esser (2005).

The results from a quantitative and a qualitative preliminary study reveal that company representatives attribute high relevance to both the initial trust before HRI and the dynamic trust during HRI with regard to the success of a cobot introduction at a workplace. In particular, the fear of loss of employment seems to be an omnipresent and existential concern among employees that can undermine trust building. In the workplace context, the latter always depends on the perceived trustworthiness of the people and institutions that carry out the development, programming, and introduction of a cobot. While there is a high level of trust in the safety of a cobot, there is a great deal of uncertainty with regard to the effects on one's own work routines and on employment security. Accordingly, production workers show a high sensitivity to all information that can be interpreted as indicators of a cobot's performance, its intended use and the human-cobot relationship. While anthropomorphism

is described as irrelevant in this context by the employees, linguistic features indicate implicit anthropomorphisation processes.

The results of an experimental study conducted in an online setting, in line with prior expectations, suggest that linguistic framing affects the perceived human-likeness of a cobot and the perceived cooperative nature of the human-cobot relationship. These two factors in turn influence initial human-cobot trust. In this context, the human-cobot relation moderates the effects of anthropomorphisation. Thus, a high degree of anthropomorphisation only affects trust towards the cobot positively under the condition that the employees situate themselves in a cooperative relationship with that cobot. The absence of this effect in case of a perceived competitive relationship illustrates the different psychological reactions to the presumed replacement by humans or machines. However, the effectiveness of framing seems to depend on the sample and/or the experimental setting, as a second experimental study reveals. This laboratory study included a real HRI addressing affective trust components and was conducted with students as the sample, who differed significantly from workers in their attitudes and prior knowledge about robots. In contrast to the first experimental study, neither did framing significantly affect students' perception of the cobot nor did the provision of relevant information about the cobot increase their trust. However, there was a strong positive effect on trust via actual interaction with the cobot. Thereby, changes in trust were not always accompanied by opposite changes in distrust.

The perceived ambiguity of the cobot seems to play a crucial role in the trust-building process and with regard to the effectiveness of linguistic framing. The cognitive complexity associated with these ambiguities due to arising mental categorization and selection problems can be resolved, following Luhmann, through trust as a commitment to a concrete expectation of the future. The mediated frames serve as discriminating factors in selecting a concrete future expectation. The individual variance of effectiveness can be explained based on individual factors such as existing prior knowledge, motivation for elaborate information processing or ambiguity aversion. These explanatory approaches require further deepening and empirical verification. Independent of framing influences, the development of trust is considered as a dynamic and multidimensional process which takes place in different phases over time and can be divided into performance- and relationship-oriented components, among others. A crucial factor influencing the development of trust is the degree of anthropomorphisation, which results in different expectations and in turn affects the dimensions of trust differently depending on the context. In addition, the effects of anthropomorphism on trust are moderated by the human-cobot relationship, which can be characterized by the tension between cooperation and competition, among other things, and has significantly more facets than suggested by the often-cited dichotomy between the representation of a robot either as a tool or as a colleague. The difference between cooperation or competition orientation can be used as a decisive characteristic of the relation at the workplace due to the widespread fear of loss of employment, but should be

supplemented in future research with further qualitative characteristics of the human-robot relationship. Thereby, a phenomenological viewpoint should be applied to investigate the employees' perception of the robot instead of addressing ontological questions about the status and actual capabilities of the robot.

Entrepreneurial and societal implications arise from the findings, which can partially be applied to other technologies and application contexts. Overall, an appropriate calibration of trust appears to be a desirable goal, which requires a careful use of language when talking about new complex technologies. The multi-method, application-oriented and interdisciplinary research approach proved to be highly suitable for the investigation of multi-layered human-robot trust, even though the empirical results are limited, for example with regard to the operationalization of the human-cobot relation or the consideration of long-term trust processes. The lack of applicability and context sensitivity of many studies has so far hampered the consolidation of existing experimental results and the development of a coherent and application-oriented theory of human-cobot trust in industrial workplaces. Thus, the latter represent overarching goals for further research. This dissertation underlines the need to integrate existing experimental results and theoretical approaches from different disciplines, such as psychology, sociology, and philosophy, in order to take sufficient account of the numerous facets, context-specific determinants and effects of trust in the workplace context.

Inhaltsverzeichnis

Kurzfassung	i
Abstract	v
Inhaltsverzeichnis	ix
Abbildungsverzeichnis	xiii
Tabellenverzeichnis	xix
Abkürzungsverzeichnis	xxiii
Danksagung	xxv
1 Einleitung	1
2 Industrielle Mensch-Roboter-Interaktion	9
2.1 Mensch-Roboter-Interaktion	9
2.1.1 Unterteilung anhand des Forschungsgegenstands	9
2.1.2 Unterteilung anhand des Interaktionstyps	11
2.2 Robotertypen	12
2.2.1 Unterscheidungskriterien	12
2.2.2 Konventionelle Industrieroboter	13
2.2.3 Kollaborationsfähige Roboter (Cobots)	15
2.2.4 Hybride Roboter	20
2.3 Industrieller Arbeitsplatzkontext	22
2.3.1 Auswirkungen auf die Arbeitswelt.....	22
2.3.2 Betriebliche Cobot-Einführungen	24
2.4 Gegenstandsbereich dieser Dissertation	30
3 Kognitive Mensch-Roboter-Interaktion	33
3.1 Mentale Modelle.....	33
3.1.1 Begriffsbestimmung	33
3.1.2 Merkmale und Funktionen	34
3.1.3 Abgrenzungen	37
3.1.4 Empirische Zugänge.....	39
3.1.5 Relevanz im MRI-Kontext	41
3.2 Anthropomorphismus	44
3.2.1 Begriffsbestimmung	44
3.2.2 Charakterisierung und Differenzierung	45
3.2.3 Theoretische Ansätze	49

3.2.4	Anthropomorphisierung von Robotern	58
3.3	Sprachliches Framing	71
3.3.1	Zur Rolle von Sprache im MRI-Kontext.....	71
3.3.2	Theoretische Ansätze	75
3.3.3	Sprachliches Framing von Robotern	79
3.4	Mensch-Roboter-Vertrauen.....	82
3.4.1	Allgemeine Merkmale und Funktionen.....	82
3.4.2	Vertrauen als mentaler Zustand	86
3.4.3	Vertrauensdefinition im MRI-Kontext.....	88
3.4.4	Kognitives und affektives Vertrauen.....	91
3.4.5	Mögliche Adressat:innen des Vertrauens im Mensch-Technik- Kontext.....	92
3.4.6	Auswirkungen auf Technologienutzung	93
3.4.7	Vertrauensmodelle	95
3.4.8	Technologieakzeptanz.....	106
3.4.9	Vertrauensmessung	110
3.4.10	Misstrauen.....	111
3.5	Zusammenführung und Forschungsprogramm.....	115
4	Empirische Arbeit	121
4.1	Methodischer Ansatz.....	121
4.2	Vorstudie 1: Online-Studie zu Erfolgsfaktoren für betriebliche Cobot- Einführungen.....	122
4.2.1	Fragestellungen und Zielsetzung.....	122
4.2.2	Identifikation und Strukturierung der Erfolgsfaktoren.....	123
4.2.3	Methoden und Material	126
4.2.4	Ergebnisse und Diskussion	129
4.3	Vorstudie 2: Qualitative Studie zu Cobot-Einführungen in der Praxis.....	135
4.3.1	Fragestellungen und Zielsetzung.....	135
4.3.2	Methoden und Material	135
4.3.3	Ergebnisse und Diskussion	140
4.4	Zwischenfazit	145
4.5	Experimentalstudie 1: Online-Studie mit Produktionsmitarbeitenden	147
4.5.1	Fragestellungen und Studiendesign.....	147
4.5.2	Methoden und Material	148
4.5.3	Ergebnisse	161
4.5.4	Diskussion.....	176
4.5.5	Zwischenfazit	185
4.6	Experimentalstudie 2: Präsenzstudie mit Studierenden.....	187
4.6.1	Fragestellungen und Studiendesign.....	187
4.6.2	Methoden und Material	188

4.6.3	Ergebnisse	196
4.6.4	Diskussion	207
4.6.5	Zwischenfazit	212
4.7	Gegenüberstellung der Experimentalstudien	213
4.7.1	Vergleich der Stichproben	213
4.7.2	Vergleich der Vertrauensentwicklungen	215
4.7.3	Einfluss des imaginierten und realen Kontakts	217
4.8	Zusammenfassung der Ergebnisse	220
5	Einbettung der Ergebnisse in die aktuelle Forschung	223
5.1	Mensch-Cobot-Vertrauen am industriellen Arbeitsplatz	223
5.1.1	Phasen der dynamischen Vertrauensentwicklung	223
5.1.2	Vertrauen als Prozess der Auflösung multipler Ambiguitäten	227
5.1.3	Vertrauen am Arbeitsplatz als multidimensionales Phänomen	230
5.1.4	Vertrauen und Misstrauen als unterschiedliche Konstrukte	233
5.2	Mensch-Roboter-Relationen und die Angst vor Arbeitsplatzverlust	236
5.2.1	Roboter als (vermeintliche) Teamkolleg:innen	236
5.2.2	Relationale Wende in der MRI	238
5.2.3	Kooperations- oder Konkurrenzorientierung als Merkmal der Relation	240
5.3	Anthropomorphismus	243
5.3.1	Impliziter Anthropomorphismus am Arbeitsplatz	243
5.3.2	Kontextabhängige Auswirkungen von Anthropomorphisierung	244
5.4	Sprachliches Framing	247
5.4.1	Prinzipielle Wirkkraft von striktem Framing	247
5.4.2	Wahrgenommene Ambiguität als Erklärungsansatz	249
5.4.3	Abhängigkeit von personenbezogenen Faktoren	251
5.4.4	Persistenz von Framing-Effekten	253
6	Implikationen, Reflexion und Ausblick	257
6.1	Implikationen	257
6.1.1	Unternehmerische Perspektive	257
6.1.2	Gesellschaftliche Perspektive	260
6.2	Übertragbarkeit der Ergebnisse	264
6.3	Methodische Reflexion	267
6.3.1	Eignung des methodischen Ansatzes	267
6.3.2	Methodische Einschränkungen	268
6.4	Ausblick auf anknüpfende Forschung	272
6.4.1	Relationale Perspektive und Anthropomorphismus	272
6.4.2	Kontextspezifische und zeitdynamische Wirkbedingungen von Framing	272

6.4.3	Replikation der Experimentalstudien	273
6.4.4	Differenzierte multidimensionale Vertrauensbetrachtung.....	273
6.4.5	Ganzheitliche und zeitdynamische Vertrauensmodelle.....	274
7	Zusammenfassung und Fazit.....	277
	Literaturverzeichnis	285
Anhang.....		333
A.	Veröffentlichungshinweise.....	333
B.	Vorstudien	335
I.	Vorstudie 1	335
II.	Vorstudie 2.....	337
C.	Experimentalstudien.....	341
I.	Framing-Text: Menschenähnlichkeit	341
II.	Framing-Text: Mensch-Cobot-Relation	342
III.	Kodierleitfaden für wahrgenommene Menschenähnlichkeit.....	344
IV.	Kodierleitfaden für wahrgenommene Kooperationsorientierung.....	345
V.	Box-Plots zur Manipulationskontrolle in Experimentalstudie 1	346
VI.	Box-Plot zur Manipulationskontrolle in Experimentalstudie 2.....	347
VII.	Übersicht Akzeptanz-Items.....	347
VIII.	Reliabilität verwendeter Skalen	349
IX.	Experimentalstudie 2: Manipulationskontrolle nach Art des Studiengangs	351
X.	Experimentalstudie 2: Konstruktionsanleitung	352

Abbildungsverzeichnis

Abbildung 2.1:	Schematische Darstellung der Forschungsgebiete der kognitiven MRI.....	10
Abbildung 2.2:	Übersicht über Interaktionsformen und deren Merkmale.	11
Abbildung 2.3:	Schematische Darstellung der Eignung verschiedener Automatisierungsvarianten in Abhängigkeit von Stückkosten und Produktionsvolumen	18
Abbildung 2.4:	Drei Beispiele für industrielle Cobots (<i>Baxter</i> von <i>Rethink Robotics</i> , <i>Yumi</i> von <i>ABB</i> , <i>UR-3</i> von <i>Universal Robots</i>)	21
Abbildung 2.5:	Phasen der Cobot-Einführung aus Sicht der betroffenen Mitarbeitenden im zeitlichen Veränderungsverlauf.....	26
Abbildung 2.6:	Bestimmung des Gegenstandsbereichs dieser Dissertation anhand vorgenannter Unterscheidungsdimensionen.....	30
Abbildung 3.1:	Dynamischer Verlauf der Intensität von Vermenschlichung in Abhängigkeit von der Interaktionsdauer.....	62
Abbildung 3.2:	Unheimlichkeitsbewertungen in Abhängigkeit von der Menschenähnlichkeit von Robotern auf Basis der <i>Anthropomorphic Robot Database</i>	70
Abbildung 3.3:	Verschiedene mögliche Konstellationen von Menschen, Robotern und Sprache.	72
Abbildung 3.4:	Theorie des überlegten Handelns.....	87
Abbildung 3.5:	Modell zum Vertrauen in Automatisierung von J. D. Lee und See (2004, S. 68).....	96
Abbildung 3.6:	Schematische Darstellung des Drei-Schichten-Vertrauensmodells nach Hoff und Bashir (2015).....	99
Abbildung 3.7:	Integriertes Vertrauensmodell.....	105
Abbildung 3.8:	Formen der Akzeptanz auf Einstellungs- und Handlungsebene (Müller-Abdelrazeq, 2020, S. 16).	107
Abbildung 3.9:	Vertrauen als Einflussvariable auf die Akzeptanz gemäß <i>automation acceptance model</i> (Ghazizadeh et al., 2012).....	108
Abbildung 3.10:	TAM-HRC für den MRI-Kontext mit empirisch ermittelten Zusammenhangsmaßen (Bröhl et al., 2019).	109
Abbildung 3.11:	Vertrauen und Misstrauen als funktional äquivalente Mechanismen zur Reduktion von Komplexität.	113

Abbildung 3.12: Zusammenspiel der theoretischen Konzepte und zentraler Forschungsansatz.	115
Abbildung 3.13: Übersicht über das Forschungsprogramm.....	119
Abbildung 4.1: Visualisierung der Prozessschritte bei der Literaturoauswahl.	124
Abbildung 4.2: Entwickeltes Rahmenwerk zur Strukturierung der Erfolgsfaktoren für eine Cobot-Einführung.....	125
Abbildung 4.3: Übersicht über Erfolgsfaktoren und deren eingeschätzte Relevanz durch Unternehmensvertreter:innen.	130
Abbildung 4.4: Anzahl der Nennung von zusätzlichen Erfolgsfaktoren in Freitextfeldern ($n = 251$).	133
Abbildung 4.5: Bedenken der Workshop-Teilnehmenden gegenüber dem Cobot-Einsatz (eigene Aufnahme).....	139
Abbildung 4.6: Assoziationen der Workshop-Teilnehmenden zur Frage nach den Bedingungen, unter welchen sie gerne mit einem Cobot zusammenarbeiten würden (eigene Aufnahme).	139
Abbildung 4.7: Schematische Darstellung des Designs von Experimentalstudie 1.....	147
Abbildung 4.8: Übersetzte Subskala der überarbeiteten <i>Godspeed</i> -Skala nach Ho und MacDorman (2017) als semantisches Differenzial zur Erhebung der wahrgenommenen Menschenähnlichkeit des Roboters.....	155
Abbildung 4.9: Eigens entwickeltes semantisches Differenzial zur Erhebung der wahrgenommenen Kooperationsorientierung.	156
Abbildung 4.10: Grafische Darstellung der wahrgenommenen Menschenähnlichkeit (<i>Godspeed</i> -Skala) je Versuchsgruppe mit inferenzstatistischen Ergebnissen.	162
Abbildung 4.11: Grafische Darstellung der wahrgenommenen Menschenähnlichkeit (Bewertung gemäß Kodierleitfaden als indirekte Messmethode) je Versuchsgruppe mit inferenzstatistischen Ergebnissen.....	163
Abbildung 4.12: Grafische Darstellung der wahrgenommenen Kooperationsorientierung der Mensch-Cobot-Relation je Versuchsgruppe mit inferenzstatistischen Ergebnissen.	163
Abbildung 4.13: Grafische Darstellung des gemessenen Vertrauensniveaus an beiden Messzeitpunkten je Versuchsgruppe.	164
Abbildung 4.14: Grafische Darstellung des gemessenen Misstrauensniveaus an beiden Messzeitpunkten je Versuchsgruppe.	165

Abbildung 4.15: Entwicklung des Vertrauensniveaus zwischen beiden Messzeitpunkten in der Experimentalgruppe (alle Versuchsbedingungen, die eine experimentelle Manipulation erhielten, gemittelt) und der Kontrollgruppe.	166
Abbildung 4.16: Interaktionsdiagramm zur Vertrauensveränderung zwischen den beiden Messzeitpunkten (mit wahrgenommener Menschenähnlichkeit abgetragen auf der horizontalen Achse).	168
Abbildung 4.17: Interaktionsdiagramm zur Vertrauensveränderung zwischen den beiden Messzeitpunkten (mit wahrgenommener Kooperationsorientierung abgetragen auf der horizontalen Achse).	169
Abbildung 4.18: Grafische Darstellung der Vertrauensveränderung zwischen beiden Messzeitpunkten je Versuchsgruppe.	169
Abbildung 4.19: Interaktionsdiagramm zur Misstrauensveränderung zwischen den beiden Messzeitpunkten (mit wahrgenommener Menschenähnlichkeit abgetragen auf der horizontalen Achse).	170
Abbildung 4.20: Interaktionsdiagramm zur Misstrauensveränderung zwischen den beiden Messzeitpunkten (mit wahrgenommener Kooperationsorientierung abgetragen auf der horizontalen Achse).	171
Abbildung 4.21: Grafische Darstellung der Misstrauensabnahme zwischen beiden Messzeitpunkten je Versuchsgruppe.	171
Abbildung 4.22: Deskriptivstatistischer Vergleich der durchschnittlichen Zustimmungswerte zu den abgefragten Akzeptanz-Items.	172
Abbildung 4.23: Schematische Darstellung des Designs von Experimentalstudie 2.	187
Abbildung 4.24: Skizze des Versuchsraums.	192
Abbildung 4.25: Arbeitsplatz und Bauteile zur Konstruktion des Getriebes.	194
Abbildung 4.26: Zeitlicher Ablauf der Interaktion.	195
Abbildung 4.27: Grafische Darstellung der wahrgenommenen Menschenähnlichkeit (<i>Godspeed</i> -Skala) je Versuchsgruppe mit Vergleich zu den Ergebnissen aus Experimentalstudie 1.	197
Abbildung 4.28: Grafische Darstellung der wahrgenommenen Menschenähnlichkeit (Bewertung gemäß Kodierleitfaden als indirekte Messmethode) je Versuchsgruppe.	198
Abbildung 4.29: Grafische Darstellung des gemessenen Vertrauensniveaus an allen drei Messzeitpunkten je Versuchsgruppe und im Vergleich zu den Versuchsteilnehmenden, bei denen Fehler in der Interaktion auftraten.	199

Abbildung 4.30: Grafische Darstellung der Vertrauensveränderung zwischen den Messzeitpunkten je Versuchsgruppe.	199
Abbildung 4.31: Grafische Darstellung des gemessenen Misstrauensniveaus an allen drei Messzeitpunkten je Versuchsgruppe und im Vergleich zu den Versuchsteilnehmenden, bei denen Fehler in der Interaktion auftraten.	200
Abbildung 4.32: Grafische Darstellung der Abnahme an Misstrauen zwischen den Messzeitpunkten je Versuchsgruppe.	201
Abbildung 4.33: Deskriptivstatistische Auswertung der Zustimmungswerte zu den abgefragten Akzeptanz-Items.	202
Abbildung 4.34: Grafische Darstellung der Zustimmungswerte zu Aussagen bezogen auf das Erleben der Interaktion, differenziert nach Untergruppe der Stichprobe mit unter- bzw. mit überdurchschnittlichem Vertrauen.	203
Abbildung 4.35: Grafische Darstellung der signifikanten Korrelationen einzelner Akzeptanzkonstrukte zu Vertrauen oder Misstrauen.	210
Abbildung 4.36: Vergleich verschiedener Einstellungen und Einschätzungen in Bezug auf Cobots von Produktionsmitarbeitenden aus Experimentalstudie 1 und Studierenden aus Experimentalstudie 2.	214
Abbildung 4.37: Vertrauensentwicklung in allen Experimentalgruppen beider Experimentalstudien über die Messzeitpunkte hinweg.	216
Abbildung 4.38: Vertrauens- und Misstrauensentwicklung zwischen den Messzeitpunkten in den jeweiligen Experimentalgruppen in Experimentalstudie 1 und Experimentalstudie 2.	217
Abbildung 4.39: Schematische Auswirkungen von imaginiertem und realem Kontakt auf Vertrauen und Misstrauen.	218
Abbildung 5.1: Mögliche dynamische Verlaufsentwicklungen des Vertrauens mit schematischen Wirkungen von Framing, realem Kontakt, passiver Inakzeptanz und Erwartungsinkongruenz.	226
Abbildung 5.2: Schaubild über multiple Ambiguitäten von Cobots, die sich auf dessen ontologischen und relationalen Status beziehen und sich in Top-Down-Schemata und Bottom-Up-Reizen manifestieren.	229
Abbildung 5.3: Gegenüberstellung der jeweiligen Verläufe von Vertrauen und Misstrauen in den beiden Experimentalstudien.	234
Abbildung 5.4: Philosophische Einordnung möglicher Fragestellungen und Forschungsfokuse im industriellen MRI-Kontext.	240

Abbildung 5.5:	Einfluss von leistungs- und relationsorientiertem Vertrauen in Abhängigkeit von der Art der Interaktion.	246
Abbildung 5.6:	Schematische Darstellung des kognitiven Prozesses bei einer tatsächlichen oder bevorstehenden Konfrontation mit einem Cobot und des Einflusses von sprachlichem Framing als <i>tie breaker</i> zur Auflösung von Ambiguität.	250
Abbildung 6.1:	Übersicht über vier mögliche Zieldomänen (Robotertypen, ähnliche Technologien, Personengruppen, Nutzungskontexte) zur Übertragung der Ergebnisse.	264
Abbildung A.1:	Verteilung der industriellen Sektoren in der Stichprobe und im gesamten deutschen verarbeitenden Gewerbe bezogen auf Beschäftigtenzahlen.	336
Abbildung A.2:	Kodierleitfaden für wahrgenommene Menschenähnlichkeit.	344
Abbildung A.3:	Kodierleitfaden für wahrgenommene Kooperationsorientierung.	345
Abbildung A.4:	Verteilung der Daten zur wahrgenommenen Menschenähnlichkeit (gemessen anhand der <i>Godspeed</i> -Skala) je Versuchsgruppe in Experimentalstudie 1 als Box-Plot.	346
Abbildung A.5:	Verteilung der Daten zur wahrgenommenen Kooperationsorientierung je Versuchsgruppe (gemessen anhand eigener Ad-hoc-Skala) in Experimentalstudie 1 als Box-Plot.	346
Abbildung A.6:	Verteilung der Daten zur wahrgenommenen Menschenähnlichkeit (gemessen anhand der <i>Godspeed</i> -Skala) je Versuchsgruppe in Experimentalstudie 2 als Box-Plot.	347
Abbildung A.7:	Konstruktionsanleitung zur Montage der Zahnradgetriebe in Experimentalstudie 2.	353

Tabellenverzeichnis

Tabelle 1:	Statistische Kennzahlen zur Durchdringung mit Industrierobotern deutschland- und weltweit.	14
Tabelle 2:	Schematischer Vergleich der typischen Merkmale von Industrierobotern und Cobots.....	16
Tabelle 3:	Funktionen von mentalen Modellen gemäß verschiedener Autor:innen.	35
Tabelle 4:	Kriterien zur Abgrenzung von mentalen Modellen gegenüber kognitiven Schemata vorwiegend nach Jones et al. (2011).....	38
Tabelle 5:	Typische Merkmale des anthropomorphen Designs.	45
Tabelle 6:	Zweistufiges Modell des Menschseins und der Vermenschlichung auf Basis von Empfindungs- und Handlungsfähigkeit nach Złotowski et al. (2017).	47
Tabelle 7:	Design-Merkmale von Robotern, die Vermenschlichung begünstigen, in Bezug zu (industriellen) Cobots.	60
Tabelle 8:	Faktoren für Vermenschlichung laut Drei-Faktoren-Theorie und deren vermutete Relevanz im Kontext der industriellen MRI.....	61
Tabelle 9:	Experimentelle Beispiele für Sozialverhalten gegenüber Robotern.	64
Tabelle 10:	Mögliche (vermeintlich) positive Auswirkungen von Anthropomorphismus.....	67
Tabelle 11:	Mögliche (vermeintlich) negative Auswirkungen von Anthropomorphismus.....	68
Tabelle 12:	Zuordnung unterschiedlicher Vertrauensdefinitionen zu den mentalen Zuständen gemäß der TRA.....	88
Tabelle 13:	Einflussfaktoren auf Vertrauen im Drei-Schichten-Modell.	97
Tabelle 14:	Vergleich der in unterschiedlichen Meta-Analysen ermittelten Effektstärken für den Einfluss bestimmter Faktorentypen auf Mensch-Roboter-Vertrauen.	102
Tabelle 15:	Verwendete Suchquellen und Suchterme bei der Literaturrecherche.....	123

Tabelle 16:	Beschriebene unternehmerische Herausforderungen und Zielsetzungen in den verwendeten Szenarien.	128
Tabelle 17:	Sprachliche Merkmale zum Framing der Menschenähnlichkeit des Cobots.....	150
Tabelle 18:	Sprachliche Merkmale zum Framing der Kooperationsorientierung der Mensch-Cobot-Relation.	151
Tabelle 19:	Länge und Lesbarkeit der verwendeten Framing-Texte in den jeweiligen Versuchsbedingungen (Experimentalstudie 1).....	152
Tabelle 20:	Items der deutschen Fassung der TAS nach Ersetzung des Begriffs <i>System</i> durch <i>Roboter</i>	153
Tabelle 21:	Textfragmente zur Abfrage des Vertrauens und Misstrauens mittels TAS jeweils vor und nach dem Framing.....	154
Tabelle 22:	Verwendete Items zur Messung verschiedener Facetten der Akzeptanz (Experimentalstudie 1).....	156
Tabelle 23:	Tabellarische Darstellung des Ablaufs von Experimentalstudie 1.....	158
Tabelle 24:	Ergebnisse der inferenzstatistischen Analyse zur Veränderung des Vertrauensniveaus zwischen beiden Messzeitpunkten je Versuchsgruppe.	164
Tabelle 25:	Ergebnisse der inferenzstatistischen Analyse zur Veränderung des Misstrauensniveaus zwischen beiden Messzeitpunkten je Versuchsgruppe.	165
Tabelle 26:	Ergebnisse der Korrelationsanalyse zwischen Akzeptanz-Items und Vertrauen bzw. Misstrauen, gemessen jeweils nach dem Framing.....	174
Tabelle 27:	Ziele am Arbeitsplatz und damit verbundene Auswirkungen in Hinblick auf das Vertrauen in einen menschenähnlich oder maschinell wahrgenommenen Cobot, zu dem eine kooperationsorientierte Relation besteht.	179
Tabelle 28:	Ziele am Arbeitsplatz und damit verbundene Auswirkungen in Hinblick auf das Vertrauen in einen menschenähnlich oder maschinell wahrgenommenen Cobot, zu dem eine konkurrenzorientierte Relation besteht.	181
Tabelle 29:	Länge und Lesbarkeit der verwendeten Framing-Texte in den jeweiligen Versuchsbedingungen (Experimentalstudie 2).....	189

Tabelle 30:	Verwendete Items zur Messung verschiedener Facetten der Akzeptanz (Experimentalstudie 2). Die Items 8 bis 13 wurden auch in Experimentalstudie 1 verwendet.	189
Tabelle 31:	Tabellarische Darstellung des Ablaufs von Experimentalstudie 2.	192
Tabelle 32:	Ergebnisse der Signifikanztests auf Unterschiede in der wahrgenommenen Menschenähnlichkeit zwischen den Versuchsgruppen, je nach Stichprobenszusammensetzung.	197
Tabelle 33:	Übersicht über gruppenspezifisch unterschiedliche Vertrauensveränderungen durch verschiedene Interventionen.	200
Tabelle 34:	Signifikante Korrelationen zwischen den Akzeptanz-Items und dem gemessenen Vertrauen sowie dem gemessenen Misstrauen, jeweils nach der Interaktion.	205
Tabelle 35:	Inferenzstatistischer Vergleich populationsbezogener Kennwerte aus Experimentalstudie 1 mit Produktionsmitarbeitenden und aus Experimentalstudie 2 mit Studierenden.	214
Tabelle 36:	Zentrale Erkenntnisse aus den verschiedenen empirischen Untersuchungen in Hinblick auf die Forschungsfragen sowie davon unabhängige explorative Erkenntnisse.	220
Tabelle 37:	Dimensionen und mögliche Ausprägungen von Vertrauen.	232
Tabelle 38:	Korrelationen zwischen Vertrauen und Misstrauen zu verschiedenen Messzeitpunkten in beiden Experimentalstudien.	234
Tabelle 39:	Items der TAS mit Zuordnung zu den Dimensionen des Vertrauens, die damit gemessen werden.	235
Tabelle 40:	Tabellarische Übersicht praktischer Implikationen aus unternehmerischer Perspektive auf Basis verschiedener wissenschaftlicher Befunde.	258
Tabelle 41:	Übersicht über Qualitätsmerkmale des Untersuchungsdesigns.	267
Tabelle 42:	Tabellarische Darstellung der in der Dissertation verwendeten eigenen Veröffentlichungen.	333
Tabelle 43:	Übersicht über die in den Experimentalstudien verwendeten Akzeptanz-Items.	347
Tabelle 44:	Reliabilitätswerte der verwendeten Skalen in Experimentalstudie 1.	349
Tabelle 45:	Reliabilitätswerte der verwendeten Skalen in Experimentalstudie 2.	350

Tabelle 46: Ergebnisse der Manipulationskontrolle für wahrgenommene Menschenähnlichkeit je nach Studiengang (technisch / nicht-technisch) der Versuchsteilnehmenden.....351

Abkürzungsverzeichnis

AAM	Automationsakzeptanzmodell (<i>automation acceptance model</i>)
ANOVA	Varianzanalyse (<i>analysis of variance</i>)
ANCOVA	Kovarianzanalyse (<i>analysis of covariance</i>)
AUC	Fläche unter der Kurve (<i>area under the curve</i>)
AV	Abhängige Variable
CASA	Computer sind soziale Akteur:innen (<i>computers are social actors</i>)
Cobot	Kollaborationsfähiger Roboter (<i>collaborative robot</i>)
fMRT	Funktionelle Magnetresonanztomographie
HN	Zur menschlichen Natur gehörend (<i>human nature</i>)
IAT	Impliziter Assoziationstest (<i>implicit association test</i>)
IDAQ	<i>Individual differences in anthropomorphism questionnaire</i> (Skala zur Messung individueller Unterschiede im Anthropomorphismus)
IFR	<i>International Federation of Robotics</i>
IMAX	<i>Images maximum</i> (Kino-System)
KI	Künstliche Intelligenz
KMU	Kleine und mittlere Unternehmen
KVP	Kontinuierlicher Verbesserungsprozess
MANOVA	Multivariate Varianzanalyse (<i>multivariate analysis of variance</i>)
MdFS	Modell der Frame-Selektion
ML	Maschinelles Lernen (<i>machine learning</i>)
MRI	Mensch-Roboter-Interaktion

MRK	Mensch-Roboter-Kollaboration
NARS	<i>Negative attitudes towards robots scale</i> (Skala zur Messung einer negativen Grundeinstellung gegenüber Robotern)
PAS	Schema der perfekten Automation (<i>perfect automation schema</i>)
SEEK	<i>Sociality, effectance and elicited agent knowledge</i> (aus der Drei-Faktoren-Theorie)
TAM	Technologieakzeptanzmodell (<i>technology acceptance model</i>)
TAM-HRC	Technologieakzeptanzmodell für MRK-Anwendungen (<i>human-robot collaboration</i>)
TAS	<i>Trust in automation scale</i> (Skala zur Messung des Vertrauens in Automationssysteme)
TPB	Theorie des geplanten Verhaltens (<i>theory of planned behavior</i>)
TRA	Theorie des überlegten Handelns (<i>theory of reasoned action</i>)
UH	Einzigartig menschlich (<i>uniquely human</i>)
UR	<i>Universal Robots</i> (Roboterhersteller)
UTAUT	<i>Unified theory of acceptance and use of technology</i> (Akzeptanzmodell)
UV	Unabhängige Variable
VR	Virtuelle Realität

Danksagung

„Alle Natur, alles Wachstum, aller Friede, alles Gedeihen und Schöne in der Welt beruht auf Geduld, braucht Zeit, braucht Stille, braucht Vertrauen.“ – Hermann Hesse

Meine Promotion zum Vertrauen in Roboter wurde erst dadurch möglich, dass sich einige Personen dazu entschieden, ihr Vertrauen in mich zu setzen. Dieses Vertrauen habe ich stets als Geschenk empfunden, dem ich mit großer Dankbarkeit begegne. Daher ist es mir ein herzliches Anliegen, all denjenigen zu danken, die mich sowohl tatkräftig und praktisch als auch emotional und mental mit ihrem Vertrauen unterstützt haben.

Zunächst gilt mein herzlicher Dank meinem Erstbetreuer Prof. Dr. Armin Grunwald, der mich während meiner Promotion wertschätzend begleitete, stets für meine Belange erreichbar war und durch sein kompetentes, inspirierendes und konstruktives Feedback maßgeblich zur Qualität meiner Arbeit beitrug. Seine Offenheit gegenüber meiner Themenwahl und dem kooperativen Promotionsverfahren sowie sein Vertrauen in meine Kompetenz haben die Anfertigung dieser Arbeit erst ermöglicht. Ebenso gilt mein herzlicher Dank meinem Zweitbetreuer Prof. Dr. Steffen Kinkel, der meine akademische Arbeit von Anfang an begleitete, von dem ich in den vergangenen Jahren viel lernen konnte und der mir die Verlässlichkeit und die Freiheit bot, mich in einem vertrauensvollen und kollegialen Arbeitsverhältnis weiterzuentwickeln und meinen Weg zu finden.

Ferner legten das Vertrauen und die Unterstützung einiger früher Förderer in meiner Schul- und Studienzeit die Grundsteine für meine akademische Laufbahn, die ich zu dieser Zeit noch nicht in Erwägung zog. Dazu zählen insbesondere Prof. Dr. Stefanie Regier, Prof. Dr. Karl Dübon, Dr. Jürgen Schöchlin und StD a. D. Erich Clemens. Sie sprachen mir Mut zu, weckten meine Neugier und boten mir wertvolle Vorbilder. Prof. Dr. Thomas Morgenstern gebührt mein herzlicher Dank für die Anregung zum Studium der Kognitionswissenschaft. Außerdem möchte ich Prof. Dr. Michael Wörz i. R. und das gesamte Team des Referats für Technik- und Wissenschaftsethik (rtwe) um Thorsten Gutsche, Sebastian Drobny und Torsten Geucke hervorheben. Mein Dank gilt ihnen für die vielfältigen Perspektiven und Lernerfahrungen, die sie mir in herzlicher Umgebung eröffneten. Die vielen inspirierenden Gedanken erwiesen sich nicht nur für die Erstellung dieser Arbeit, sondern für meinen gesamten Lebensweg als bereichernd.

Herzlicher Dank gebührt ferner meinen geschätzten ehemaligen und aktuellen Kolleg:innen am Institut für Lernen und Innovation in Netzwerken (ILIN) für die konstruktive Zusammenarbeit. Insbesondere bedanke ich mich bei meinem langjährigen Weggefährten Dennis Richter, der stets ein offenes Ohr für all meine Sorgen im Arbeitsalltag und darüber

hinaus hatte. Unsere heiß umkämpften Duelle am Kicker boten den notwendigen Ausgleich und bleiben unvergessen. Ganz besonderer Dank gebührt auch Marco Baumgartner für die vielen intensiven und gewinnbringenden Diskussionen, sein unerschöpfliches Vertrauen in die Qualität meiner Arbeit und seine ebenso schier endlose Geduld, meine kritischen und mitunter zweifelnden Nachfragen zu ertragen. Die stets reibungslose Zusammenarbeit trug maßgeblich zum Gelingen der Arbeit bei. Mein Dank gilt ebenso Prof. Dr. Brita Schemmann, die mir auch nach unserer gemeinsamen Zeit am ILIN stets als zuverlässige und kompetente Ratgeberin zur Seite stand.

Zudem bedanke ich mich bei den Beteiligten des Projektes ProBot, das den geeigneten Rahmen für meine anwendungsnahe empirische Forschung bot. Insbesondere Dr. Tobias Kunkel unterstützte mich mit freundschaftlichem und konstruktivem Rat weit über das übliche Maß hinaus. Christian Bauer danke ich für das überaus professionelle und gewinnbringende Lektorat, Dr. Bettina-Johanna Krings, Lena Kölmel und Lina Kluy für ihre wertvollen Rückmeldungen zu ausgewählten Kapiteln meiner Dissertation sowie Stephan Kühner-Sobiesinsky für die Unterstützung bei all den gestalterischen Fragestellungen. Außerdem danke ich meinen Kolleg:innen am Institut und allen Freund:innen, die sich bereitwillig und teils kurzfristig als Korrekturleser:innen, Diskussionspartner:innen und Zuhörer:innen bereitstellten, allen voran Dr. Carla Schmidt, Tobias Zundel, Miriam Spieß, Daniel Rothschnitt, Christian Bauer, Nadjila Nasar, Sarah Kuhn, Viktoria Steenhuis und Karin Ginzky. Für die finanzielle Unterstützung im Rahmen des erhaltenen Promotionsstipendiums danke ich der Heinrich-Hertz-Gesellschaft am Karlsruher Institut für Technologie (KIT).

Zuletzt möchte ich meinem privaten Umfeld für die unschätzbar wertvolle Unterstützung während der Promotionszeit und auf meinem Lebensweg danken. Mein aufrichtiger Dank gilt zunächst meinen Eltern. Mit Fleiß, Ausdauer und Disziplin gaben sie mir die notwendigen Tugenden für dieses Projekt mit auf meinen Lebensweg und ließen nie einen Zweifel daran, mich auf diesem zu unterstützen und mir mit aller Selbstverständlichkeit den Rücken freizuhalten. Gleichmaßen danke ich allen Freund:innen, Weggefähr:innen und Unterstützer:innen für ihre treue Begleitung, für ihren herzlichen Zuspruch und kompetenten Beistand in schwierigen Phasen. Diese Unterstützung trug dazu bei, dass ich die notwendige innere Stärke, das Selbstvertrauen und das Durchhaltevermögen für ein solches Projekt gewinnen konnte. Von ganzem Herzen danke ich dir, Julchen, für deine aufrichtige Zuneigung, Treue und tiefe Verbundenheit, für den stetigen Beistand im gemeinsamen Alltag, dein Verständnis für manche missmutigen Tage voller Zweifel und dein Vertrauen in mich und meine Fähigkeiten. Es war mir ein Geschenk!

Wörth am Rhein, im Juli 2022

Tobias Kopp

1 Einleitung

Ob als therapeutische Roboter-Robbe *Paro*, als künstliches Haustier *Aibo* (Becker et al., 2013, S. 55ff.) oder gar als Sexroboter zur Befriedigung intimer Bedürfnisse (Bendel, 2019b; Levy, 2017) – moderne Roboter fungieren in vielfältiger Form als künstliche Begleiter:innen (*artificial companions*) und kreieren neue Qualitäten der Mensch-Roboter-Beziehung (Böhle & Bopp, 2014). Mithin wird selbst die These vertreten, dass sie liebende Ehepartner:innen repräsentieren können (Levy, 2017). Bedingt durch technologische Neuerungen im Bereich der Robotik entwickelt sich eine Bandbreite komplexerer Beziehungsformen jenseits der lange Zeit vorherrschenden hierarchischen Sichtweise des Roboters als bloßes Werkzeug des Menschen. Roboter werden zu einem „relationalen Artefakt“ (*relational artifact*)¹ und „Begleiter“ (*companion*) (Turkle, 2002, S. 150). Damit einhergehende Fragestellungen nach der adäquaten Beschreibung aktueller Mensch-Technik-Verhältnisse und damit verbundener Hierarchien werden seit jeher verhandelt, gewinnen aber angesichts technischer Errungenschaften eine neue Aktualität, zumal die heutigen Relationen als qualitativ neue Dimension empfunden werden (Heßler, 2019). Daran knüpft sich die Frage nach der Selbstdeutung von Menschen in Bezug und in der Abgrenzung zu Maschinen, z. B. als defizitärer Gegenpart, als Gestalter:in, als Kontrolleur:in oder als Kollaborateur:in, wobei die verschiedenen Konzeptionen durchaus koexistieren (Heßler, 2019, S. 53). Diese verschiedenen Deutungen finden sich in gängigen Metaphern und Narrativen wieder und manifestieren die Perspektive auf Technologie und auf Roboter im Speziellen (Canellas et al., 2017, S. 15). Dabei beschreibt, interpretiert und konstituiert Sprache die Mensch-Roboter-Relationen und lässt den Roboter mitunter als Quasi-Gegenüber (*quasi-other*) innerhalb einer sozialen Beziehung erscheinen (Coeckelbergh, 2010b, 2011, S. 61). Das Mensch-Roboter-Vertrauen als zentrales Charakteristikum einer solchen Relation gewinnt dabei an Relevanz und Facettenreichtum, zumal kontextspezifisch höchst unterschiedliche Dimensionen des Vertrauens relevant sind und der Prozess der Vertrauensbildung verschiedenste Betroffene einschließt (J. D. Lee & See, 2004). Nicht zuletzt diskutieren Forschende die Rückwirkungen auf Mensch-Mensch-Beziehungen, die aufgrund ihrer vergleichsweise hohen Komplexität im Kontrast zu zunehmend vertrauten Roboterbeziehungen als mühsam empfunden werden könnten (Turkle, 2011).

¹ Zum Zweck des besseren Leseflusses sind in dieser Dissertation wörtliche Zitate aus englischsprachiger Fachliteratur im Text ins Deutsche übersetzt. Häufig werden nachfolgend die englischsprachigen Originalbegriffe in Klammern und kursiv gesetzt angeführt. Dabei handelt es sich um einschlägige Fachbegriffe oder besonders eingängige Ausdrücke aus der englischsprachigen Originalliteratur, die sich nur unzureichend ins Deutsche übersetzen lassen und wörtlich übernommen werden.

Noch tauchen derartige Roboter in Form künstlicher Begleiter:innen isoliert in spezifischen Anwendungssettings wie bspw. im Rahmen der Pflege dementer Personen auf und haben kaum Einzug in die alltägliche Lebenswelt der breiten Bevölkerung gehalten. Dennoch werden sie in den Medien und in der Forschung häufig thematisiert. Hierbei besteht die Gefahr, dass ein Missverhältnis zwischen ambitionierten Zukunftsvisionen und gegenwärtigen Anwendungsmöglichkeiten dazu führt, dass notwendige Debatten über drängende aktuelle Fragen in Bezug auf bestehende Technologien, deren Gestaltung und Nutzung ausbleiben oder zumindest in den Hintergrund rücken, wie Kehl und Coenen (2016) in Bezug auf Anwendungen der künstlichen Intelligenz (KI) feststellen. Daher fokussiert die vorliegende Dissertation explizit eine aktuell existierende Technologie und deren Nutzungskontext, nämlich industriell verwendete Roboter. Diese stellen einen geeigneten Untersuchungsgegenstand dar, da sie bereits in großer Zahl im praktischen Einsatz sind und den Arbeitsalltag vieler Menschen beeinflussen. Insofern stellt „die Industriehalle de[n] paradigmatische[n] Herkunfts- und Einsatzort real existierender Roboter“ (Remmers, 2020, S. 56) neben dem wachsenden Bereich der Service-Robotik dar.

Industrieroboter gelten seit Jahrzehnten als Erfolgsfaktor in der industriellen Fertigung (Stadler, Weiss, Mirnig & Tscheligi, 2013, S. 232) und prägen dort das Bild wie kaum eine andere Technologie (Moniz & Krings, 2016). Diese Roboter agieren in Bereichen, die durch physische Schutzvorrichtungen strikt vom Arbeitsbereich der Mitarbeitenden getrennt sind und damit keine Interaktion zulassen (Breazeal, 2004; Heyer, 2010; May et al., 2017). Produktionsprozesse mit hohen Flexibilitätsanforderungen oder hoher Komplexität lassen sich allerdings bis dato nicht kosteneffizient vollautomatisieren, sondern erfordern die Kompetenzen menschlicher Mitarbeitender (Lenz, 2011, S. 12). Folgerichtig zeichnet sich in der deutschen Wirtschaft in den letzten Jahren ein zunehmendes Interesse an sog. kollaborationsfähigen Robotern (Cobots) ab, die eine direkte, physische Mensch-Roboter-Interaktion (MRI) in einer gemeinsamen Arbeitsumgebung ermöglichen (Campa, 2016; Charalambous, Fletcher & Webb, 2016b; Weiss, Huber, Minichberger & Ikeda, 2016). Durch ihre höhere Wandlungsfähigkeit eignen sich solche Roboter besser für die Produktionsbedingungen abseits der klassischen Serienproduktion (Steil & Maier, 2017a, S. 3) und sind damit insbesondere für kleine und mittlere Unternehmen (KMU) attraktiv (Bloss, 2016). Expert:innen erwarten, dass die Arbeitswelt der Zukunft von der Erweiterung menschlicher Fähigkeiten durch Kollaboration mit Maschinen geprägt sein wird (Campolo, Sanfilippo, Whittaker & Crawford, 2017, S. 8). Je mehr und je intensiver Mitarbeitende mit diesen Robotern interagieren, desto spürbarer wandelt sich ihr Arbeitsalltag und ihr Selbstverständnis in Relation zur Technologie (Meißner, Trübswetter, Conti-Kufner & Schmidtler, 2020). Während die technische Realisation der MRI seit einigen Jahrzehnten intensiv beforscht wird, bestehen gegenwärtig allerdings noch keine umfänglichen Theorien auf Seite der kognitiven MRI, die sich aus einer menschenzentrierten Perspektive deziert mit der Wahrnehmung und mentalen Repräsentation der Interaktion sowie der Interaktionspartner:innen beschäftigt.

Während die Fragen nach der Qualität komplexer werdender Mensch-Roboter-Interaktionen und -Relationen im Bereich der sozialen Robotik und der Service-Robotik verstärkt gestellt werden, bleiben diese im Kontext des industriellen Arbeitsplatzes weitgehend unberücksichtigt. Während soziale Roboter im Sinne einer intuitiven Interaktion darauf programmiert sind, gezielt soziale Verhaltensmuster zu imitieren und Gefühle bei ihren menschlichen Gegenübern auszulösen (Breazeal, Dautenhahn & Kanda, 2016; Darling, 2016), steht bei industriellen Robotern traditionell der funktionale Mehrwert im Sinne eines hilfreichen Werkzeugs im Fokus. Doch selbst in diesem Anwendungskontext gerät die Umschreibung des heutigen Zusammenspiels aus Menschen und Robotern durch die klassischen und simplen Mensch-Werkzeug-Relationen an ihre Grenzen. Vielmehr avancieren Roboter zunehmend zu Teammitgliedern (Roesler & Onnasch, 2020b) und wirken damit potenziell auf die Arbeitszufriedenheit von Mitarbeitenden, die maßgeblich davon abhängt, ob diese sich als geschätzten Bestandteil eines Teams fühlen (Nyholm & Smids, 2020). Das Verschwimmen der Grenzen zwischen sozialer Robotik und weiter gefasster MRI-Forschung (Richert, Müller, Schröder & Jeschke, 2017) gebietet es, die strikte Unterscheidung zwischen dem Roboter als Werkzeug und dem Roboter als sozialem Gegenüber auf den Prüfstein zu stellen (Breazeal, 2004; Moniz & Krings, 2016). Moderne Cobots sind zwar funktional für den Einsatz im industriellen Kontext gestaltet, lassen sich aber durch ihr steigendes Maß an Interaktionsfähigkeit (Darling, 2017) und durch die mitunter vorhandenen anthropomorphen Gestaltungsmerkmalen wie simulierte menschliche Augen oder eine menschenähnliche Morphologie bspw. mit *Roboterarmen* (Elprama, El Makrini, Vanderborght & Jacobs, 2016, S. 1) kaum mehr einer einzelnen Roboterklasse zuordnen. Sie gelten daher als hybride Roboter (Brandstetter, 2017, S. 18f.), welche „die Grenzen zwischen Werkzeuggebrauch und Zusammenarbeit langsam auflösen“ (Remmers, 2020, S. 57).

Mitarbeitende, die mit solchen hybriden Robotern interagieren, sind mit deren Mehrdeutigkeit konfrontiert und nehmen dementsprechend ambivalente Haltungen ein. Diese bilden den schmalen Grat zwischen den jeweils naheliegenden, aber konfligierenden mentalen Repräsentationen eines Cobots entweder als soziales, autonom agierendes Wesen oder als typisches mechanistisches Werkzeug ab (Hinds, Roberts & Jones, 2004; Weiss et al., 2016). Aus dieser Vieldeutigkeit erwächst die Herausforderung, diese Roboter mit unseren gegenwärtigen sprachlichen Mitteln zu fassen und in die bewährten Denkkategorien einzuordnen. Die Sprachverwendung stößt insofern an ihre Grenzen, als dass kein Vokabular zur Verfügung steht, um über technologische Artefakte zu sprechen, die mehr als nur Dinge sind (*more than things*). Das provoziert eine Spannung zwischen den affektiven Wahrnehmungen und Gefühlen auf der einen Seite und der sozial akzeptierten, sprachlich fassbaren und tradierten Sprechweise auf Basis eines Subjekt-Objekt-Dualismus auf der anderen Seite (Coeckelbergh, 2018). Trotz einer gewissen Forschungstradition gilt daher die MRI weiterhin als ein exploratives Forschungsfeld (Broadbent, 2017). Insbesondere das komplexe Verhältnis zwischen Sprache und Technologie bedarf einer verstärkten empirischen

Untersuchung unter Berücksichtigung technikphilosophischer Theorien (Coeckelbergh, 2017).

Die industrielle MRI stellt damit aus dem Blickwinkel der kognitiven MRI ein relevantes und attraktives anwendungsnahe Forschungsfeld dar, das in der bisherigen Forschung unterrepräsentiert ist (Charalambous, Fletcher & Webb, 2017; Meißner & Trübswetter, 2018, S. 225). Traditionell dominieren bspw. Fragen der Arbeitssicherheit oder der Ergonomie die Forschungslandschaft im Vergleich zu menschbezogenen Aspekten (*human factors*), bei denen die Erfahrungswelt der Mitarbeitenden im Fokus steht (Charalambous, Fletcher & Webb, 2016a; Charles, Charalambous & Fletcher, 2015; Müller, Shehadeh, Schröder, Richert & Jeschke, 2018; Obrist, Reitberger, Wurhofer, Förster & Tscheligi, 2011, S. 145f.; Sheridan, 2016; Vincent, Taipale, Sapio, Lugano & Fortunati, 2015, S. 2). Ferner werden Fragen der Akzeptanz im Gegensatz zum Bereich der Service-Robotik am industriellen Arbeitsplatz zumeist kaum mitgedacht (Savela, Turja & Oksanen, 2018), obwohl eine unzureichende Benutzungsfreundlichkeit im Speziellen und eine mangelnde Akzeptanz der Mitarbeitenden im Allgemeinen in der Praxis als zentrale Barrieren für eine flächendeckende Verbreitung von Cobots gelten (Weiss et al., 2016, S. 2). Die Annahme, dass Mitarbeitende im industriellen Bereich schlichtweg zur Nutzung verpflichtet werden können, greift zu kurz und vernachlässigt den negativen Einfluss innerer Widerstände gegenüber der Technologienutzung auf die Zufriedenheit der Mitarbeitenden einerseits sowie auf die Effizienz der Interaktion andererseits (Ullrich, Vladova, Thim & Gronau, 2015). Die Einführung von Cobots als hybride Roboter unterstreicht den Bedarf nach Studien, die psychologische und soziale Faktoren im industriellen Kontext adressieren (Bröhl, Nelles, Brandl, Mertens & Schlick, 2017; El Makrini, Elprama, van den Bergh & Lefebvre, 2016; Müller, Schröder, Jeschke & Richert, 2017, S. 448; Müller-Abdelrazeq, 2020; Siciliano & Khatib, 2016, S. 1416; Steil & Maier, 2017a, S. 9, 2017b, S. 404; Vanderborgh et al., 2017; Weiss & Huber, 2016). Seit wenigen Jahren steigt ferner das Bewusstsein für die Notwendigkeit, soziale Aspekte und Implikationen der MRI am Arbeitsplatz stärker interdisziplinär zu erforschen (Moniz & Krings, 2016) und Cobots dabei als sozio-technische Systeme zu konzeptionieren (Decker, 2017; Frennert, 2019, S. 311; Moniz, 2014, S. 127).

Mensch-Roboter-Relationen lassen sich nicht zuletzt durch das Vertrauen des Menschen in Bezug auf den Roboter charakterisieren (Charalambous et al., 2016b; Moniz & Krings, 2016; Phillips, Ososky, Grove & Jentsch, 2011). Dabei steht der Roboter als vermeintliche:r Adressat:in des Vertrauens nur stellvertretend für dahinterstehende Personen, Institutionen und Systeme aus dem Organisationskontext, deren (wahrgenommene) Vertrauenswürdigkeit für den Vertrauensbildungsprozess entscheidend ist.² Die in der öffentlichen Debatte häufig einseitig positive Konnotation des Mensch-Technik-Vertrauens (Sumpf,

² Die Rede vom Vertrauen in einen Roboter im Rahmen dieser Einleitung ist daher als vereinfachend anzusehen. Eine genauere Analyse der tatsächlichen Vertrauensnehmer:innen findet sich in Kapitel 3.4.5.

2019, S. 90) als zu maximierende, graduelle Größe (Fletcher & Webb, 2017, S. 167) über-
sieht dabei dessen Komplexität und Facettenreichtum. So kann ungerechtfertigtes Ver-
trauen zu fälschlichen Erwartungen führen und sich damit nachteilig auf eine MRI auswir-
ken (Hancock, Kessler, Kaplan, Brill & Szalma, 2020; Parasuraman & Riley, 1997).
Speziell im Bereich der industriellen Arbeit stellt die Angst vor der Ersetzung durch eine
Technologie und vor dem Verlust des eigenen Arbeitsplatzes ein Spezifikum der jeweili-
gen Mensch-Roboter-Relation und eine entscheidende Facette des Vertrauens von Mitar-
beitenden dar. Die Relevanz dieses Faktors ist gemeinhin anerkannt (Elprama, Jewell,
Jacobs, El Makrini & Vanderborght, 2017; Richert et al., 2017; Salvini, Laschi & Dario,
2010), zumal er mit der Aussicht auf eine sinnstiftende Arbeit einen zentralen Aspekt des
menschlichen Daseins tangiert (Nyholm & Smids, 2020). Dennoch wurde die Angst vor
Arbeitsplatzverlust als Einflussfaktor auf Vertrauen bisher nur unzureichend in Studien zur
kognitiven MRI einbezogen und untersucht (Granulo, Fuchs & Puntoni, 2019).

Einen weiteren Einflussfaktor auf das Vertrauen stellt der Grad der Vermenschlichung
(Anthropomorphisierung) des Roboters dar. Hierbei handelt es sich um einen psychischen
Prozess, der dazu führt, dass Personen unbelebten Entitäten mentale Zustände zuschreiben,
die gemäß dem *common sense* Menschen vorbehalten sind (Bergamasco & Herr, 2016,
S. 1900f.; Compagna, Weidemann, Marquardt & Graf, 2016; Darling, 2016, 2017; Phillips
et al., 2011). Selbst die Aktionen mechanisch simpel gestalteter, abstrakter, nicht-anthro-
pomorpher Roboter können sozial-emotionale Deutungsmuster und Reaktionen hervorrufen
(Erel, Hoffman & Zuckerman, 2018). (Hybride) Roboter befördern diese menschliche
Tendenz aufgrund ihrer spezifischen Charakteristika zusätzlich (Marquardt, 2017). Dies
schlägt sich auch sprachlich bspw. in Bezeichnungen wie „hybride[] Teams‘ oder ‚Robo-
ter[] als Kollegen“ (Onnasch, Jürgensohn, Remmers & Asmuth, 2019, S. 33) nieder. Mit-
hilfe der Theorie der mentalen Modelle lassen sich Vermenschlichungsvorgänge als ent-
lehrende Übertragung bestehender gedanklicher Strukturen begreifen. Solche mentalen
Modelle dienen als komplexitätsreduzierende interne Repräsentation des Roboters, um
dessen zukünftige Aktionen, Reaktionen und Interaktionen zu antizipieren (Johnson-Laird,
2005; Jones, Ross, Lynam, Perez & Leitch, 2011; Seel, 2017, S. 934). Gerade in Erman-
gelung realer Erfahrungswerte im Umgang mit Cobots kann es vorkommen, dass Mitar-
beitende Repräsentationen von Menschen auf Roboter übertragen (S. Lee, Kiesler, Lau &
Chiu, 2005; Phillips et al., 2011, S. 1491). Ein hohes Maß an Vermenschlichung wird dabei
gemeinhin als vertrauensfördernd betrachtet (Waytz, Heafner & Epley, 2014), wenngleich
uneinheitliche empirische Ergebnisse einen starken moderierenden Einfluss von Kontext-
variablen nahelegen (Ötting, Masjutin, Steil & Maier, 2020).

Ob es im konkreten Fall zu einer Vermenschlichung eines Roboters kommt, hängt abgese-
hen von roboterbezogenen Faktoren wie dessen Erscheinungsbild und von intrapersonalen
Faktoren wie der Persönlichkeit des interagierenden Menschen (Bergamasco & Herr, 2016,
S. 1901) vom Kontext der MRI ab (Darling, 2017). Dazu gehört u. a. das sog. sprachliche

Umfeld (*linguistic environment*) der Interaktionssituation (Coeckelbergh, 2011, S. 65). Der Begriff des sprachlichen Framings umschreibt hierbei die Möglichkeit, mittels gezielter Formulierungen bestimmte Erwartungs- und Deutungsrahmen zu aktivieren, innerhalb derer konkrete Mensch-Roboter-Interaktionen verstanden und interpretiert werden (Groom et al., 2011; Kory & Kleinberger, 2014; Stenzel, Chinellato, del Pobil, Lappe & Liepelt, 2012; Stenzel, Chinellato, Bou et al., 2012). Experimentelle Studien zeigen, dass die Bezeichnung von Robotern mit menschlichen Namen anstelle technischer Termini oder das Hinzufügen einer Hintergrundgeschichte die mentale Repräsentation des Roboters, das Vertrauen in denselben und dementsprechend die konkrete Wahrnehmung einer Interaktion in einem Top-Down-Verarbeitungsprozess beeinflussen (Darling, 2017, S. 11; S. Lee et al., 2005). Allerdings fokussierte sich die Forschung lange Zeit ausschließlich auf den Einfluss von Bottom-Up-Verarbeitungsprozessen, die durch die experimentelle Manipulation sichtbarer Merkmale initiiert wurden (Coeckelbergh, 2011; Groom et al., 2011). Die empirische Evidenz zur Auswirkung von Framing im Kontext realer MRI-Szenarien am Arbeitsplatz und insbesondere zu deren Nachhaltigkeit ist noch lückenhaft (Kory & Kleinberger, 2014; Kory Westlund, Martinez, Archie, Das & Breazeal, 2016; Onnasch & Roesler, 2019).

Die Zielsetzung dieser Dissertation besteht daher darin, das Vertrauen in Cobots mit seinen vielfältigen Facetten und praktischen Auswirkungen sowie dessen Beeinflussbarkeit durch sprachliches Framing am industriellen Arbeitsplatz unter Einbezug der dort vorherrschenden Mensch-Cobot-Relationen empirisch zu untersuchen. Damit trägt das Forschungsvorhaben zur Ergründung der bisher unzureichend erforschten menschbezogenen Faktoren bei der Interaktion mit hybriden Cobots am Arbeitsplatz bei. Ferner sollen die Forschungsergebnisse ein profunderes und empirisch untermauertes Verständnis der relevanten (kognitions-)psychologischen Konstrukte und deren Zusammenspiel ermöglichen und somit die Ausbildung einer umfassenden, kontextsensitiven und praktisch anwendbaren Theorie zur industriellen MRI unterstützen. Überdies liefert die Forschungsarbeit weiterführende Einblicke über die Rolle der Sprache in Bezug auf Technikvertrauen. An dieser Stelle eröffnen sich Anknüpfungspunkte für philosophische und soziologische Fragestellungen, die sich zumindest teilweise auf Mensch-Roboter-Relationen in anderen Kontexten oder die Wahrnehmung anderer moderner Technologien wie KI-Anwendungen übertragen lassen. Aus praktischer Sicht können die gewonnenen Erkenntnisse dazu beitragen, eine nachhaltig gelingende Mensch-Cobot-Interaktion zu ermöglichen und damit verbundene volkswirtschaftliche Potenziale zu erschließen. Aus methodischer Sicht verfolgt das Forschungsvorhaben einen anwendungsnahen, methodenpluralistischen und interdisziplinären Ansatz. Der Einsatz quantitativer sowie qualitativer Untersuchungsmethoden sowie der Einbezug theoretischer Ansätze aus dem Bereich der Psychologie, Kognitionswissenschaft, Philosophie, Soziologie und Wirtschaftswissenschaft verspricht einen tiefgehenden und ganzheitlichen Blick auf das komplexe Phänomen. Die relevanten psychologischen Konstrukte werden im Sinne einer hohen Reliabilität möglichst über validierte

Selbstauskunftsskalen erhoben, ergänzt durch implizite Messmethoden. Die Untersuchung einer real stattfindenden MRI, der Rückgriff auf Produktionsmitarbeitende als Stichprobe und der explizite Einbezug der typischen Spezifika des Arbeitsplatzkontexts, allen voran der Angst vor Arbeitsplatzverlust, begünstigen die externe Validität.

Die Dissertation gliedert sich im Wesentlichen in einen theoretischen, einen empirischen und einen zusammenführenden Teil. In den folgenden Kapiteln 2 und 3 werden die für diese Arbeit relevanten theoretischen Grundlagen zur Betrachtung von Mensch-Roboter-Interaktionen erläutert. Die Unterteilung in zwei Kapitel folgt der Differenz zwischen industrieller und kognitiver MRI. Kapitel 2 behandelt dezidiert die industrielle MRI, also entsprechende Interaktionen, die im industriellen Umfeld stattfinden, nimmt Differenzierungen der Robotertypen und Interaktionskontexte vor und umreißt damit den Gegenstandsbereich der vorliegenden Dissertation. Die in Kapitel 3 behandelte kognitive MRI umfasst die relevanten theoretischen Konstrukte, die zum Verständnis der menschlichen Wahrnehmungen und Reaktionen in Bezug auf eine MRI notwendig sind. Die beschriebenen Themenkomplexe werden aus einer primär kognitionswissenschaftlichen Perspektive beleuchtet und – wenn immer dies produktiv erscheint – um dezidiert philosophische und soziologische Ansätze ergänzt. Dadurch werden gemäß dem angestrebten interdisziplinären Charakter stets produktive Anschlüsse zu angrenzenden Forschungsdisziplinen geschaffen. Im abschließenden Unterkapitel 3.5 wird das Zusammenspiel der thematisierten Themenfelder dargestellt und daraus das Forschungsprogramm der vorliegenden Dissertation abgeleitet. Kapitel 4 umfasst die komplette empirische Arbeit, die sich entlang einer quantitativen und einer qualitativen Vorstudie sowie anschließend entlang zweier primär quantitativer Experimentalstudien strukturiert. Letztere bauen auf den Erkenntnissen aus den Vorstudien auf. Dabei beantworten die Studien nach und nach die Forschungsfragen. Im abschließenden Unterkapitel 4.8 werden die empirischen Erkenntnisse tabellarisch in Bezug zu den behandelten Forschungsfragen zusammengefasst. Kapitel 5 bettet die empirischen Ergebnisse aus sämtlichen Studien in die dargestellten theoretischen Konzepte und die aktuelle Forschungsliteratur ein und vertieft dieselbe. In Kapitel 6 erfolgt im Rahmen einer Abschlussdiskussion eine Erläuterung der sich ergebenden praktischen und gesellschaftlichen Implikationen. Ferner findet sich in diesem Kapitel eine Analyse der Übertragbarkeit der Ergebnisse sowie eine kritische Reflexion der eingesetzten Methodik und eine Skizzierung attraktiver Anknüpfungspunkte für weitere Forschungsarbeiten. Das abschließende Kapitel 7 fasst die Ergebnisse der Arbeit kompakt zusammen.

2 Industrielle Mensch-Roboter-Interaktion

Dieses Kapitel widmet sich der industriellen Mensch-Roboter-Interaktion. Kapitel 2.1 differenziert die MRI zuerst anhand zweier Unterscheidungskriterien in verschiedene Subdisziplinen. Einerseits erfolgt eine Unterteilung anhand des jeweiligen Forschungsgegenstands in physische und kognitive MRI, andererseits anhand des jeweiligen Interaktionstyps in Koexistenz, Kooperation und Kollaboration. Anschließend werden in Kapitel 2.2 verschiedene Robotertypen klassifiziert, die im industriellen Umfeld Anwendung finden. Kapitel 2.3 fokussiert die Spezifika des industriellen Arbeitskontexts und beschreibt die Phasen betrieblicher Cobot-Einführungsprozesse sowie die dabei typischen Wahrnehmungen und Reaktionen der Mitarbeitenden. Auf Basis der vorgenommenen Differenzierungen wird schließlich in Kapitel 2.4 der Gegenstandsbereich der vorliegenden Dissertation konkretisiert und eingegrenzt.

2.1 Mensch-Roboter-Interaktion

2.1.1 Unterteilung anhand des Forschungsgegenstands

Innerhalb der MRI-Forschung haben sich unterschiedliche Strömungen herausgebildet, die Mensch-Roboter-Interaktionen aus diversen Blickwinkeln betrachten und dabei verschiedene Forschungsgegenstände in den Fokus rücken. Anhand dieses fokussierten Forschungsgegenstands lässt sich die MRI-Forschung in die beiden Subdisziplinen der physischen MRI und der kognitiven MRI unterteilen (Mutlu, Roy & Šabanović, 2016; Xing & Marwala, 2018, S. 4). Diese noch junge Unterscheidung wird bisher zwar nur in wenigen Arbeiten explizit vorgenommen, erweist sich allerdings insbesondere im Kontext der vorliegenden Dissertation als gewinnbringend, um die behandelten Themen von rein technischen Fragestellungen abzugrenzen.

Die physische MRI befasst sich aus dem Blickwinkel der Ingenieurwissenschaften mit der Entwicklung von interaktiven Robotern und deren notwendigen Funktionen, um sichere, effiziente und intuitive physische Interaktionen mit Menschen zu ermöglichen. Im Gegensatz dazu fokussiert die kognitive MRI die kognitiven Fähigkeiten und Konzepte. Innerhalb der Forschung zur kognitiven MRI werden menschliche Interaktionspartner:innen, der beteiligte Roboter und deren gemeinsame Aktionen als kognitives Gesamtsystem betrachtet.

Letzteres lässt sich anhand dreier Arten von Modellen analysieren, die den zentralen Forschungsgegenstand der kognitiven MRI bilden. Erstens repräsentieren Menschen den Roboter mittels mentaler Modelle (vgl. Kapitel 3.1). Zweitens erlauben implementierte Modelle dem Roboter eine Interaktion mit der Umgebung. Drittens lässt sich die Interaktion, Kommunikation und Koordination zwischen Menschen und Robotern durch konzeptionelle Modelle abbilden (Mutlu et al., 2016; Xing & Marwala, 2018, S. 7). Abbildung 2.1 visualisiert schematisch die unterschiedlichen Modelle.¹

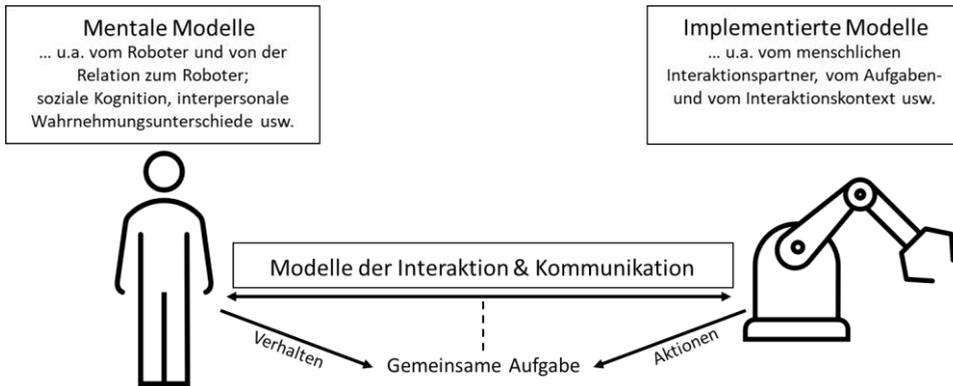


Abbildung 2.1: Schematische Darstellung der Forschungsgebiete der kognitiven MRI, übersetzt und leicht adaptiert von Mutlu et al., 2016, S. 1908.

Ein zentrales Ziel der kognitiven MRI besteht in einem tieferen Verständnis der menschlichen Wahrnehmungen, Erwartungen und Reaktionen in Bezug auf Roboter im Rahmen einer Interaktion (Mutlu et al., 2016). Im Gegensatz zur physischen MRI erfolgt die Untersuchung dieser Gegenstände primär aus dem Blickwinkel der Kognitionswissenschaften und/oder der (Industrie-)Psychologie. Die Forschung zur kognitiven MRI hat zwar in den vergangenen Jahren an Bedeutung gewonnen, gilt allerdings als vergleichsweise unterrepräsentiert. Mit der Verfügbarkeit interaktiver Roboter, die eine wesentlich engere und vielseitigere Zusammenarbeit ermöglichen und zunehmend in den menschlichen Nahbereich vordringen, nimmt die Komplexität und Relevanz der menschzentrierten Fragestellungen zu. Damit erhöht sich auch der Bedarf an Studien zur Technologieakzeptanz im industriellen Kontext, die dezidiert psychologische und soziale Faktoren einbeziehen (Bröhl et al., 2017; El Makrini et al., 2016; Müller et al., 2017; Müller-Abdelrazeq, 2020,

¹ Zur Sprachregelung: Ein mentales Modell, das ein Mensch ausbildet, um einen Roboter damit zu repräsentieren, wird im Folgenden als mentales Modell *vom* Roboter bezeichnet. Die im Modell repräsentierte Instanz (das *modellandum*) steht als Objekt hinter dem Modell-Begriff, wohingegen das Adjektiv vor dem Modell-Begriff anzeigt, um wessen Modell es sich handelt. Nähere Erläuterungen zu Modellen finden sich in Kapitel 3.1.

1f.; Steil & Maier, 2017a, S. 9, 2017b, S. 404; Vanderborght et al., 2017; Weiss & Huber, 2016). Die Erkenntnisse der kognitiven MRI können die physische MRI wiederum informieren und zu einer nutzungsorientierten Roboterentwicklung beitragen, sodass moderne Roboter möglichst effizient mit ihrer Umwelt agieren und sich dabei kongruent zu den mentalen Modellen der Interaktionspartner:innen verhalten (Mutlu et al., 2016, S. 1908).

2.1.2 Unterteilung anhand des Interaktionstyps

Hinweis: Einige Inhalte aus diesem Kapitel (inklusive der Grafiken) sind in Kopp, Schäfer und Kinkel (2020) veröffentlicht.

Die Interaktion zwischen Menschen und Robotern lässt sich auf verschiedene Arten realisieren, die sich hinsichtlich der Intensität der Interaktion unterscheiden. Die häufig synonyme Verwendung von MRI und Mensch-Roboter-Kollaboration (MRK) ist dabei irreführend (Bender, Braun, Rally & Scholtz, 2016), denn MRI umfasst als Überbegriff verschiedenartige Interaktionsformen, darunter u. a. die MRK (A. Bauer, Wollherr & Buss, 2007; Onnasch, Maier & Jürgensohn, 2016). Bislang fehlt es an einer einheitlichen Taxonomie, um mögliche Typen der MRI zu klassifizieren (Aaltonen, Salmi & Marstio, 2018; Onnasch et al., 2016). Bestehende Ansätze unterteilen Interaktionen in drei bis sechs Unterkategorien (Aaltonen et al., 2018; Bender et al., 2016; Malik & Bilberg, 2019; Onnasch et al., 2016). Abbildung 2.2 stellt die wichtigsten Interaktionstypen und deren Merkmale dar.

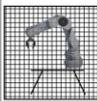
	Vollautomatisierung mit Industrierobotern			Mensch-Roboter-Interaktion mit Cobots	
	Kapselung		Koexistenz	Kooperation	Mensch-Roboter-Kollaboration
					
Arbeitsschritte	sequenziell			gleichzeitig	
Arbeitsraum	getrennte Arbeitsräume			gemeinsamer Kollaborationsraum falls synchronisiert: zeitliche Trennung	
Arbeitsaufgabe	keine Kopplung der Aufgaben			gekoppelte Aufgaben	gemeinsame Aufgabe
Physischer Kontakt	nicht möglich			möglich, nicht notwendig	möglich, häufig erwünscht
Mindestanforderungen gem. DIN EN ISO 10218-1	Automatikbetrieb mit trennender Schutzeinrichtung	sicherheitsbewerteter überwachter Halt	Geschwindigkeits- und Abstandsüberwachung Leistungs- und Kraftbegrenzung Handführung		
Geschwindigkeit	maximale Geschwindigkeit		begrenzte Geschwindigkeit		

Abbildung 2.2: Übersicht über Interaktionsformen und deren Merkmale.

Bei der herkömmlichen Vollautomatisierung sind die Arbeitsräume der Mitarbeitenden und der Roboter durch physische Schutzvorrichtungen voneinander getrennt

(Roboterzelle), sodass keine direkte Interaktion möglich ist. Im Gegensatz dazu kennzeichnen sich sämtliche Formen der Interaktion dadurch, dass keine physischen Separierungen zwischen den Mitarbeitenden und den Robotern existieren. Bei der Koexistenz teilen sich die beteiligten Menschen und Roboter dennoch keinen gemeinsamen Arbeitsraum. Begegnungen entstehen nur sporadisch und der Roboter stoppt seine Arbeit, sobald er einen Menschen in seinem Arbeitsraum detektiert (Dieber, Schlotzhauer & Brandstötter, 2017; Onnasch et al., 2016). Ein physischer Kontakt ist nur im Betriebsstopp-Modus möglich (Aaltonen & Salmi, 2019). Die Interaktion beschränkt sich demnach auf die Vermeidung von Kollisionen, also explizit auf das Verhindern physischer Interaktion. Dennoch erfordert diese Vermeidungsstrategie ein Monitoring der Umgebung und eine interaktive Verhaltensanpassung an etwaige Aktionen des jeweils anderen. Bei der Kooperation führen die beteiligten Menschen und Roboter einzelne aufeinander folgende Arbeitsschritte sequenziell und im Regelfall ohne physischen Kontakt innerhalb eines gemeinsamen Arbeitsraums durch. Bei der sog. synchronisierten Kooperation teilen sich die Menschen zwar einen Arbeitsraum mit dem Roboter, operieren darin allerdings nie zur gleichen Zeit. Die intensivste Stufe der Zusammenarbeit stellt die Kollaboration dar, bei der die Mitarbeitenden mit dem Roboter im selben Arbeitsraum, dem sog. Kollaborationsraum, gleichzeitig eine gemeinsame Aufgabe verrichten und dabei mitunter am selben Werkstück arbeiten (Aaltonen & Salmi, 2019). Durch die mangelnde Trennschärfe der Interaktionstypen und die Pluralität an Definitionsansätzen lassen sich in der Praxis insbesondere kooperative und kollaborierende Szenarien nur schwer unterscheiden. Sämtliche Interaktionsformen erfordern mit speziellen Sicherheitsfunktionen ausgestattete kollaborationsfähige Roboter (Cobots; vgl. Kapitel 2.2.3). Dieser Robotertypus steht im Fokus der vorliegenden Dissertation. Im folgenden Kapitel findet eine Einordnung der bestehenden Robotertypen und eine Abgrenzung zu konventionellen Industrierobotern statt.

2.2 Robotertypen

2.2.1 Unterscheidungskriterien

Der Begriff des Roboters fand erstmals im tschechischen Drama *R.U.R. Rossum's Universal Robots* (Čapek, 1920/2012) Verwendung. Etymologisch lässt sich der Begriff des Roboters als Abwandlung des tschechischen Begriffs *robotá* für (Zwangs-)Arbeit und Fronddienst bestimmen (Loh, 2019, S. 16; Müller-Abdelrazeq, 2020, S. 7). Bei Robotern handelt es sich um eine Unterklasse der Maschinen, die durch ihre Programmierbarkeit in der Lage sind, eine Vielfalt an Aktionen jeweils repetitiv auszuführen. Im Gegensatz zu rein virtuellen Agenten, die auf Computern mittels Algorithmen simuliert werden, besitzen Roboter stets eine physische Gestalt (Müller-Abdelrazeq, 2020, S. 8f.). Diese physische Verkörperung ist auch eines von sechs Kriterien, die Misselhorn (2013, S. 43) als allgemeine

Definition für Roboter vorschlägt. Ferner nennt sie als Kriterien, dass ein Roboter über einen Prozessor verfügt, um Informationen zu verarbeiten, Sensoren besitzt, um Umwelteinflüsse aufzunehmen, sowie über Effektoren oder Aktoren mechanische Abläufe initiieren und physisch auf seine Umgebung einwirken kann. Zudem handelt ein Roboter autonom bzw. erweckt zumindest den Anschein, dies zu tun (zitiert nach Loh, 2019, S. 16f.).

In den letzten Jahrzehnten wurden verschiedene Typen von Robotern entwickelt, die sich anhand unterschiedlicher Kriterien wie ihrem äußeren Erscheinungsbild (maschinell, anthropomorph, humanoid, animaloid), ihrer Fähigkeiten (z. B. soziale Roboter) oder ihrem typischen Anwendungskontext klassifizieren lassen. Die Unterscheidung nach dem Anwendungskontext in Industrie- und Serviceroboter findet sich in der grundlegenden ISO-Norm über „Roboter und Robotergeräte“ (ISO 8373:2012) wieder. Dieser Definition folgt auch die *International Federation of Robotics* (IFR). Während Industrieroboter innerhalb des industriellen Produktionsprozesses eingesetzt werden und bspw. Werkstücke bearbeiten oder logistische Aufgaben übernehmen, erbringen Serviceroboter nützliche Dienstleistungen für Menschen, wie bspw. die Säuberung des Haushalts durch Staubsaugerroboter (International Federation of Robotics [IFR], 2020a, 2020b).

2.2.2 Konventionelle Industrieroboter

ISO 8373:2012 definiert einen Industrieroboter als „automatisch gesteuerte[n], frei programmierbare[n] Mehrzweck-Manipulator, der in drei oder mehr Achsen programmierbar ist und zur Verwendung in der Automatisierungstechnik entweder an einem festen Ort oder beweglich angeordnet sein kann“ (zitiert nach Wischniewski, Rosen & Kirchhoff, 2019, S. 2). Solche Roboter werden in der Industrie bereits seit Jahrzehnten eingesetzt, um routinemäßige Arbeitsschritte mit hoher Wiederholungsrate schnell und akkurat durchzuführen. Dabei operieren sie in aller Regel in statischen Produktionsumgebungen zum Zweck der Fertigung großer Serien, die nur in seltenen Fällen eine Umrüstung und Umprogrammierung des Roboters erfordern (Heyer, 2010; Vanderborght et al., 2017). Aus Sicherheitsgründen werden Industrieroboter in eingehausten Zellen betrieben, da sie mit hoher Geschwindigkeit und großer Kraftwirkung operieren und damit ein erhebliches Verletzungsrisiko für Menschen darstellen (Buchner, Wurhofer, Weiss & Tscheligi, 2012; Fletcher & Webb, 2017; Heyer, 2010; Vanderborght et al., 2017; Wilcox, Nikolaidis & Shah, 2013). Analog zu anderen Maschinen in der industriellen Fertigung gelten solche Roboter zumeist als bloße Werkzeuge des Menschen (Broadbent, 2017, S. 628).

Die industrielle Fertigung gilt als eines der ersten realen Anwendungsgebiete für Roboter im Produktivbetrieb (May et al., 2017, S. 165). Die Kennzahlen in Tabelle 1 verdeutlichen die merkliche Zunahme an Industrierobotern in diesem Sektor. Deutschland bildet dabei den weitaus größten Absatzmarkt für Industrieroboter in Europa und gehört neben China,

Japan, den USA und Südkorea zu den fünf Staaten, die gemeinsam fast drei Viertel der weltweiten Roboterinstallationen (74 %) vereinen. Insbesondere der deutsche Automobilsektor spielt eine entscheidende Rolle. Dieser weist eine der weltweit höchsten Durchdringungsraten auf und hat einen Anteil von 50 % an allen in Deutschland eingesetzten Industrierobotern (International Federation of Robotics, 2019). Gleichwohl variiert der Durchdringungsgrad je nach Tätigkeitsbereich. Die Karosseriefertigung in der Automobilindustrie erfolgt laut Expert:innenschätzungen bereits zu 90 % automatisiert durch Industrieroboter, wohingegen dieser Wert in der Montage mit ca. 20 % deutlich geringer ausfällt (Steil & Maier, 2017a, S. 2). Im Brüsseler *Audi*-Werk befinden sich gar von 550 im Betrieb befindlichen Industrierobotern nur sechs in der Montagelinie, obwohl die hohen Lohnkosten und der Fachkräftemangel auch in diesem Fertigungsbereich die Frage nach zusätzlichen Automatisierungsmöglichkeiten aufwerfen (Vanderborghet et al., 2017).

Tabelle 1: Statistische Kennzahlen zur Durchdringung mit Industrierobotern deutschland- und weltweit. Datenquelle: International Federation of Robotics (2019).

	Deutschlandweit	Weltweit
Neuinstallationen von Industrierobotern in 2018	26.723	422.271
Durchschnittliche Veränderung der jährlichen Neuinstallationen im Zeitraum 2013-2018	+8 % p.a.	+19 % p.a.
Im Betrieb befindliche Roboter in 2018	ca. 215.800	ca. 2.5 Mio.
Durchschnittliche Veränderung der im Betrieb befindlichen Industrieroboter im Zeitraum 2013-2018	+5 % p.a.	+13 % p.a.
Roboterdichte im produzierenden Gewerbe pro 10.000 Mitarbeitende	338 (Automobilsektor: 1.268)	99

Trotz dieser bemerkenswerten Wachstumswahlen bestehen technische und organisatorische Grenzen für den Einsatz von Industrierobotern. Die notwendige Flexibilität zur Fertigung verschiedener Produktvarianten und das Handling geringfügiger Materialunterschiede gelten als Herausforderungen, die sich mithilfe von Industrierobotern derzeit nicht kosteneffizient bewältigen lassen (Charalambous et al., 2016b, S. 193). So sah sich bspw. das Unternehmen *Toyota* veranlasst, den Robotereinsatz an einigen Stellen wieder zurückzufahren (Der Tagesspiegel, 2019). Als Gründe wurden hohe Kosten, häufige Stillstände,

mangelhaftes Handling materialbedingt leicht unterschiedlicher Werkstücke und eine ausbleibende Weiterentwicklung des Produktionssystems genannt. Diese Weiterentwicklung lebte zuvor von Vorschlägen der Mitarbeitenden im Rahmen eines kontinuierlichen Verbesserungsprozesses (KVP). An zunehmender Popularität gewann daher in den vergangenen Jahren der Ansatz, Roboter eng mit Mitarbeitenden zusammenarbeiten zu lassen (Buchner et al., 2012; Campa, 2016; Onnasch et al., 2016), um damit dem Trend hin zu einer flexibleren Produktion und der Fertigung kundenindividueller Produkte (*mass customization / mass personalization*) gerecht zu werden (Shehadeh, Schröder, Richert & Jeschke, 2017, S. 1208; Wang, Ma, Yang & Wang, 2017). Das artikulierte Ziel besteht dabei in der Verknüpfung der jeweiligen Stärken von Menschen (z. B. Fingerfertigkeit, kognitive Fähigkeiten) und Robotern (z. B. Kraft, Ausdauer, Wiederholgenauigkeit) zu einem leistungsfähigeren und effizienteren Gesamtsystem (Lenz, 2011, S. 3; Robla-Gomez et al., 2017). In diesem Zusammenhang werden verstärkt kollaborationsfähige Roboter (Cobots) als technische Lösungen diskutiert.

2.2.3 Kollaborationsfähige Roboter (Cobots)

2.2.3.1 Definition und Abgrenzung

Das Kofferwort *Cobot* setzt sich aus den englischen Begriffen *collaborative* und *robot* zusammen. Die US-amerikanischen Professoren Peshkin und Colgate prägten und patentierten das Konzept der Cobots (Akella et al., 1999; Peshkin & Colgate, 1999a). In der Patentschrift verweisen sie explizit auf den Aspekt der direkten physischen Interaktion zwischen einer Person und dem Roboter (Peshkin & Colgate, 1999b), der sich auch in der heutigen ISO-Definition 8373:2012 wiederfindet (Murashov, Hearl & Howard, 2016, S. 6). Die kollaborierende Arbeitsweise stellt entsprechend das zentrale Merkmal von Cobots dar, welches sich in deren Bezeichnung widerspiegelt und diese von konventionellen Industrierobotern abgrenzt.

Der Begriff des Cobots wird häufig unscharf sowohl für solche Roboter verwendet, die tatsächlich direkt mit einem menschlichen Gegenüber kollaborierend eine Aufgabe bearbeiten, als auch für klassische Leichtbauroboter, denen die entsprechenden Sicherheitsvorkehrungen für den kollaborierenden Einsatz fehlen (Schou, Andersen, Chrysostomou, Bøgh & Madsen, 2018, S. 73). Auch die IFR differenziert in statistischen Daten nicht, ob ein Cobot tatsächlich kollaborierend eingesetzt wird. Ein Roboter gilt vielmehr dann als Cobot, wenn er herstellerseitig mit den entsprechenden Funktionen zur kollaborierenden Nutzung ausgestattet ist (International Federation of Robotics, 2019, S. 54). Daher wird in der vorliegenden Dissertation der Begriff des kollaborationsfähigen Roboters eingeführt. Dieser umfasst all jene Roboter, die aufgrund ihrer technischen Ausstattung die gefahrlose Kollaboration mit einem Menschen ermöglichen, ungeachtet dessen, ob dieser Roboter in der realen praktischen Verwendung tatsächlich kollaborierend eingesetzt wird. Die

Umschreibung als kollaborationsfähige Roboter erscheint insofern treffender und eindeutiger, als sie auf den Möglichkeitsraum an Einsatzszenarien verweist. Kollaborationsfähige Roboter können stets auch im Rahmen weniger intensiver Interaktionen wie der Kooperation, der Koexistenz oder der Kapselung verwendet werden.

Cobots sind für den Einsatz in der Produktion konzipiert. Entsprechend gilt der industrielle Sektor in der nahen Zukunft als mit Abstand wichtigstes Anwendungsfeld für kollaborationsfähige Roboter (Steil & Maier, 2017b, S. 404). Somit lassen sich Cobots den Industrierobotern zuordnen. Im Rahmen der vorliegenden Dissertation bezieht sich zur besseren Abgrenzung der Begriff des (klassischen oder konventionellen) Industrieroboters dezidiert auf solche Roboter, die aufgrund fehlender Sicherheitsfunktionen ohne direkte physische Interaktion mit dem Menschen betrieben werden. Tabelle 2 stellt die typischen Unterschiede zwischen Cobots und klassischen Industrierobotern dar. Dabei ist zu beachten, dass diese schematische Auflistung die Komplexität verkürzt, die durch die verschiedenartige Ausgestaltung der Merkmale bei konkreten Robotermodellen entsteht. So investieren bspw. zwischenzeitlich auch Hersteller von Industrierobotern in die Reduktion der Programmierkomplexität (Kopp, Hendig & Kinkel, 2021).

Tabelle 2: Schematischer Vergleich der typischen Merkmale von Industrierobotern und Cobots (stark angelehnt an Djuric, Urbanic & Rickli, 2016, S. 458).

	Industrieroboter	Cobot
Installationsort	festinstalliert mit physischer Einhausung	relativ leicht änderbar; ggf. auf mobiler Plattform
Anwendungsgebiet	variantenarme Produktion, große Serien, seltene Umrüstungen	variantenreiche Produktion, kleine Serien, häufige Umrüstungen
Komplexität der Programmierung	zumeist hohe Komplexität, spezielle Programmierkenntnisse erforderlich	geringe Komplexität, bspw. durch einfache Methoden des Einlernens durch Handführung
Grad der Interaktion	keine Interaktion möglich	alle MRI-Typen inkl. Kollaboration möglich
Sicherheitsmechanismen	Trennung durch Schutzzaun	gemeinsamer Arbeitsraum, spezielle Sicherheitsfunktionen implementiert
Physikalische Maße	mittelgroß bis groß	klein
Arbeitsgeschwindigkeit	hoch	niedrig

Mit der Entwicklung von Cobots im Rahmen der Roboterforschung und schließlich mit der kommerziellen Verfügbarkeit einiger Modelle Mitte der 2010er-Jahre gewinnt die MRK-Forschung zunehmend an Relevanz (Steil & Maier, 2017a, S. 2). Durch das Verschwinden von Schutzzäunen entsteht ein gemeinsamer Arbeitsraum für die Mitarbeitenden und den Roboter. Der direkte Kontakt mit dem Cobot eröffnet die Möglichkeit für einfachere Mechanismen der Programmierung, etwa über das sog. kinästhetische Einlernen, also das Führen des Roboterarms (Malik & Bilberg, 2019). Dadurch eignen sich Cobots im Vergleich zu Industrierobotern für dynamischere Produktionsumgebungen mit einer größeren Produktvielfalt, kleineren Serien und diversen Einzelarbeitsschritten (Schou & Madsen, 2017), die für zunehmend relevantere und durch Kundenbedürfnisse getriebene Produktionsformen wie *mass customization* oder *mass personalization* relevant sind (Decker, Fischer & Ott, 2017, S. 348). Ferner versprechen Cobots Automatisierungsmöglichkeiten in denjenigen Produktionsbereichen, in denen Industrieroboter bisher kaum eingesetzt werden, z. B. in der Montage, wie eine Untersuchung von Bøgh, Hvilshøj, Kristiansen und Madsen (2012) zeigt. Laut deren Analyse von 566 einzelnen Montageaufgaben in fünf Produktionsstätten eines großen Unternehmens könnten etwa zwei Drittel davon in den kommenden Jahren mithilfe von Cobots automatisierbar sein, wohingegen nur 13 % selbst langfristig nicht umsetzbar erscheinen. Dabei eignen sich insbesondere logistische Aufgaben wie Transport und Teilezuführung sowie Assistenzaufgaben wie die (Vor-)Montage für die Automatisierung durch Cobots. Serviceorientierte Aufgaben wie Wartung und Säuberung erscheinen erst in ferner Zukunft realisierbar.

2.2.3.2 Ökonomische Betrachtungen

Hinweis: Einige Inhalte aus diesem Kapitel (einschließlich der Abbildung) sind in Kopp et al. (2020) veröffentlicht.

Aus betriebswirtschaftlicher Sicht besetzt die MRK eine Nische zwischen rein manueller Arbeit und roboterbasierter Vollautomatisierung, wie Abbildung 2.3 schematisch darstellt. Bevor die MRK als Produktionsform zur Verfügung stand, war es bei einem Produktionsvolumen $> v_{\text{man, Roboter}}$ ökonomisch sinnvoll, manuelle Arbeit direkt durch die roboterbasierte Automatisierung abzulösen. Mit der Einführung der MRK ist es nun bereits bei einem vergleichsweise kleineren Produktionsvolumen $> v_{\text{man, Cobot}}$ ökonomisch ratsam, die bisher rein manuelle Arbeit um Cobots zu erweitern, also eine MRK-Lösung zu etablieren. Dabei ist zu bedenken, dass die menschliche Arbeitskraft bei der MRK im Gegensatz zur roboterbasierten Vollautomatisierung weiterhin direkt zur Wertschöpfung beiträgt. Erst ab der Schwelle $v_{\text{Cobot, Roboter}}$ ist es ökonomisch geboten, auf konventionelle Industrieroboter zu setzen. Diese weisen im Normalbetrieb bei hohen Stückzahlen ohne Umrüstvorgänge eine höhere Performance im Vergleich zu Cobots auf, deren Arbeitsgeschwindigkeit aus Sicherheitsgründen begrenzt ist. Durch die Einführung der MRK hat sich der Schwellenwert für das Produktionsvolumen, ab dem konventionelle Roboter die wirtschaftlichste

Alternative darstellen, nach oben verschoben ($V_{\text{Cobot, Roboter}} > V_{\text{man, Roboter}}$). Damit vergrößert sich der ökonomisch sinnvolle Einsatzbereich menschlicher Arbeitskraft, sei es in Form rein manueller Arbeit oder in Zusammenarbeit mit einem Cobot (Steil & Maier, 2017b).

Einschränkend sei erwähnt, dass die tatsächlichen Kurvenverläufe von weiteren Faktoren abhängen und einer schematischen *ceteris paribus*-Betrachtung entsprechen. Zu diesen Faktoren gehören u. a. die perspektivisch sinkenden Anschaffungskosten sowohl für konventionelle als auch für kollaborationsfähige Robotik, die bereits bei kleineren Produktionsvolumina zu einer stärkeren Fixkostendegression führen. Ferner bleiben bei dieser Betrachtungsweise nicht-monetäre Anreize zur Einführung von Automatisierungstechnologien wie die hohe Wiederholgenauigkeit oder die Attraktivierung der Arbeitsplätze genauso unberücksichtigt wie stufenartige Anstiege in den Kurven bei gewissen Produktionsvolumina, die die Anschaffung zusätzlicher Robotereinheiten erfordern.

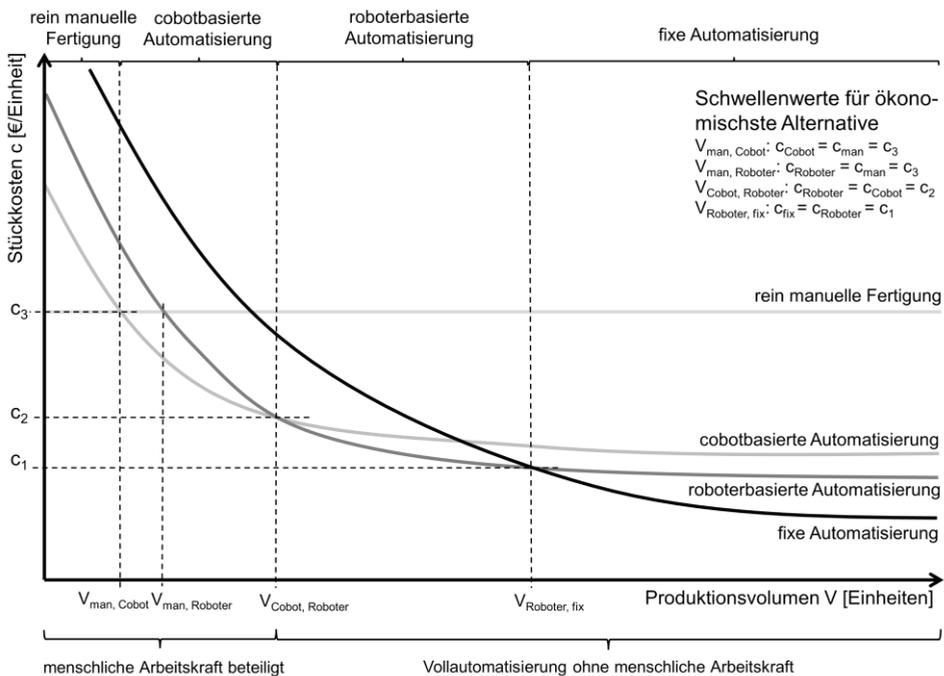


Abbildung 2.3: Schematische Darstellung der Eignung verschiedener Automatisierungsvarianten in Abhängigkeit von Stückkosten und Produktionsvolumen (in Anlehnung an Matthias & Ding, 2013).

2.2.3.3 Gegenwärtige und perspektivische Marktsituation

Die Robotisierung stellte bisher insbesondere für Großunternehmen eine wirtschaftlich attraktive Möglichkeit zur Automatisierung von Produktionsprozessen dar. Mit der

Verfügbarkeit von Cobots wird die industrielle Verwendung von Robotern erstmals auch für viele KMU ökonomisch sinnvoll (Moniz, Fischer & Krings, 2017, S. 8). Entsprechend schreiben Expert:innen Cobots ein besonders hohes Marktpotenzial in KMU zu (Görke et al., 2017). Das gilt perspektivisch auch für die Übernahme von Nicht-Routine-Arbeiten (Decker et al., 2017, S. 349), falls Roboter künftig dank Methoden der KI und des maschinellen Lernens (ML) flexibler agieren und sich selbstständig an menschliche Arbeitsweisen anpassen können (Shehadeh et al., 2017, S. 1208). Der Megatrend zur durchgängigen Digitalisierung von Produktionsprozessen vereinfacht dabei die Integration der Cobots zusätzlich (Steil & Maier, 2017b, S. 405f.).

Entsprechend verwundert es wenig, dass in den letzten fünf bis zehn Jahren geäußerte Marktprognosen überwiegend sehr positiv ausfielen. So wurden Cobots als „benötigte einwandfreie Mitarbeitende“ (*the flawless co-workers needed*) (Richert, Shehadeh, Müller, Schröder & Jeschke, 2016, S. 1) bezeichnet, die die Arbeitsweise in KMU drastisch verändern werden (Sauppé & Mutlu, 2015). Forschende vermeldeten ein hohes Interesse aus der Industrie (Charalambous et al., 2016a) und einen an Fahrt gewinnenden Einsatz in KMU (El Makrini et al., 2018). Mitunter wurde gar von einer baldigen und weit verbreiteten Einführung von Cobots im Rahmen einer dritten Welle der Robotisierung nach der Einführung klassischer Industrieroboter in der Produktion in den 1970er-Jahren und der Einführung von Service-Robotern während der Jahrtausendwende ausgegangen (Fletcher & Webb, 2017, S. 159; Murashov et al., 2016, S. 2). Diesen insgesamt optimistischen Prognosen stehen in den letzten Jahren skeptischere Veröffentlichungen gegenüber, die eine weniger dynamische Einführung vermuten lassen. Unternehmen scheinen das den Cobots zugeschriebene Potenzial bis dato nur zu einem geringen Grad zu erkennen (Baumgartner, Kopp & Kinkel, 2020) und zu realisieren (Cherubini, Passama, Crosnier, Lasnier & Fraisse, 2016; Dieber et al., 2017; Görke et al., 2017; Oubari, Pischke, Jenny, Meißner & Trübswetter, 2018). Hierzu trägt auch ein Mangel an unabhängigen empirisch validierten Erkenntnissen zur tatsächlichen Wirtschaftlichkeit und zu erzielten Effizienzsteigerungen bei (Hegenberg, Schimpf, Fischer & Schmidt, 2019). Das veranlasste die IFR im Jahr 2019 zu der Vermutung, dass die Prognosen für das weltweite Marktwachstum von Cobots überbewertet (*over-hyped*) seien (Litzenberger, 2019). Trotz eines erheblichen Wachstums der Cobot-Verkaufszahlen von 2017 auf 2018 um 23 % bleiben Cobots mit einem Anteil von unter 4 % (Stand: 2018) an allen installierten Robotern eine Randerscheinung (International Federation of Robotics, 2019, S. 54). Insbesondere mit einer zeitnahen flächendeckenden Verbreitung kollaborierender Anwendungsfälle in den meisten produzierenden Industrien sei daher nicht zu rechnen (International Federation of Robotics, 2018, S. 2).

Für die schleppende Einführung von Cobots ist neben hohen Sicherheitsanforderungen und einer damit verbundenen geringen Flexibilität in der betrieblichen Verwendung auch die hohe Unsicherheit bezüglich des Umgangs mit den interaktiven Fähigkeiten

verantwortlich, die inzwischen mittels praxisbezogener Forschung adressiert wird (Steil & Maier, 2017b, S. 405f.). Die restriktiven Sicherheitsanforderungen ergeben sich aus der entsprechenden Norm über die „Sicherheitsanforderungen für Robotersysteme im industriellen Umfeld“ (DIN EN ISO 10218-1). Eine normgerechte und gleichzeitig effiziente MRI galt lange Zeit als kaum umsetzbar. Erst seit Erscheinen der technischen Spezifikation ISO/TS 15066 (DIN EN ISO/IEC 27001), die genauere Richtwerte für den Betrieb von Cobots festlegte, und durch entsprechende Weiterentwicklungen der Cobots hat sich diese Situation verbessert. Inzwischen liegt ferner eine Novellierung der DIN EN ISO 10218 im Entwurfsstadium vor (Stand: April 2021).

Zur Erfüllung der Sicherheitsvorgaben greifen die Roboterhersteller auf unterschiedliche technologische Ansätze zurück, die von verschiedenartigen Kraft- und Kollisionssensoren bis hin zu passiven elastischen Elementen reichen (Malik & Bilberg, 2019; Steil & Maier, 2017b, S. 405). Trotz dieser implementierten Sicherheitsvorkehrungen besteht die Pflicht zur Durchführung einer Risikobeurteilung des kompletten Cobot-Arbeitsplatzes. Diese mitunter aufwändige Prozedur stellt ein zentrales Hemmnis für den Cobot-Einsatz dar, da dadurch der Spielraum für flexible Adaptionen verengt und der zentrale Vorteil der Flexibilität im Einsatz gemindert oder gar unterminiert wird. Ferner mangelt es vielen Firmen an Erfahrungswerten bezüglich der Spezifikation und der Kalkulation interaktiver Robotersysteme. Die Einrichtung des gesamten Cobot-Arbeitsplatzes, einschließlich dessen Werkzeugausstattung, sensorischer Konfiguration und der entsprechenden Peripherie, stellt entsprechend eine zusätzliche Herausforderung für Unternehmen dar (Schou et al., 2018, S. 72). Abgesehen von diesen produktionstechnischen Barrieren gilt die mangelnde Akzeptanz der Mitarbeitenden als einflussreiche Barriere für Cobot-Einführungen (You & Robert, 2018, S. 251). Um die Bedenken der Mitarbeitenden besser zu verstehen, ist es wichtig, Cobots als hybride Roboter zu begreifen (vgl. Kapitel 2.2.4) und die spezifischen Rahmenbedingungen der Technologieverwendung am industriellen Arbeitsplatz zu untersuchen (vgl. Kapitel 2.3).

2.2.4 Hybride Roboter

Hybride Roboter lassen sich weder eindeutig der Klasse der sozialen Roboter noch der Industrieroboter zuordnen, da sie in ihrer äußerlichen Gestaltung und in ihren Funktionalitäten sowohl Merkmale klassischer funktionaler als auch sozialer Roboter aufweisen (Weiss et al., 2016). Der Begriff des hybriden Roboters trägt damit dem Umstand Rechnung, dass die Grenzen zwischen funktional gestalteten, zunehmend interaktiven Industrierobotern und sozialen Service-Robotern verschwimmen und neue Roboterklassifizierungen erforderlich werden (Brandstetter, 2017, S. 18f.). So argumentieren bspw. Tanner, Burkhard & Schulze (2019, S. 1), dass sich auch Industrieroboter unter gewissen Umständen der Gruppe der sozialen Roboter zuordnen lassen, sofern sie z. B. über kommunikative

Fähigkeiten verfügen. Müller et al. (2017, S. 448) konstatieren umgekehrt, dass soziale Roboter zunehmend in Produktionsumgebungen eingesetzt werden. Laut Richert et al. (2017, S. 75) wurde der Begriff des sozialen Roboters um die Jahrtausendwende u. a. durch Billard und Dautenhahn (1997) geprägt, ohne dass sich seitdem eine eindeutige Definition herauskristallisierte. In verschiedenen Publikationen werden unterschiedliche Aspekte dezidiert hervorgehoben, um soziale Roboter zu charakterisieren. Gemein ist den Definitionen, dass die Roboter auf einer emotionalen Ebene agieren (Darling, 2016, S. 4), sich an von Menschen erwarteten sozialen Verhaltensnormen orientieren (Bartneck & Forlizzi, 2004; Richert, 2018) und damit besonders zu ihrer Vermenschlichung anregen (Phillips et al., 2011, S. 1493). Letzteres kann durch eine humanoide bzw. anthropomorphe Robotergestaltung unterstützt werden, die überdies die soziale Bindung zu dem Roboter begünstigt (Groom, Takayama, Ochi & Nass, 2009, S. 2) und philosophische sowie anthropologische Fragen nach der *conditio humana* neu stellt (Ferrari & Eyssel, 2016, S. 909). Gemeinhin lassen sich soziale Roboter gegenüber industriellen oder Service-Robotern dadurch abgrenzen, dass sie dazu konzipiert sind, menschliche Gefühle hervorzurufen oder Sozialverhalten nachzuahmen (Darling, 2016). Allerdings schließen sich die Ausstattung mit Funktionen zur Erledigung industrieller Aufgaben sowie mit sozialen Mechanismen keineswegs wechselseitig aus.

Hybride Roboter erscheinen menschlichen Rezipienten mitunter als eine ambivalente Wesensart, die sich kaum in bestehende mentale Modelle (vgl. Kapitel 3.1) integrieren lässt. Einerseits handelt es sich *de facto* um Maschinen, andererseits ähneln solche Roboter durch ihr (simuliertes) Sozialverhalten eher Lebewesen als Maschinen und lösen entsprechende soziale Reaktionen aus (Ferrari & Eyssel, 2016, S. 912; Kory Westlund et al., 2017). Im Gegensatz zu konventionellen Industrierobotern werden hybride Roboter mitunter als soziale Entitäten oder Kooperationspartner:innen wahrgenommen, was die Integration in die Arbeitsabläufe und in das soziale Gefüge am Arbeitsplatz erleichtern kann (Sauppé & Mutlu, 2015).



Abbildung 2.4: Drei Beispiele für industrielle Cobots (*Baxter* von *Rethink Robotics*, *Yumi* von *ABB*, *UR-3* von *Universal Robots*), Bildquelle: Phillips, Zhao, Ullman & Malle (2018).

Abbildung 2.4 visualisiert anhand dreier gängiger Cobots das Spektrum der anthropomorphen Gestaltung industrieller Cobots. Zwar sind weiterhin die meisten Cobot-Modelle maschinenähnlich gestaltet und ähneln in ihrem äußeren Erscheinungsbild klassischen Industrierobotern (Elprama et al., 2016), dennoch weisen Cobots Merkmale hybrider Roboter auf. Dazu zählen ihre geringe Größe, ihre häufig menschenähnliche Morphologie mit abgerundeten Gelenken, die Fähigkeit zur direkten Interaktion einschließlich physischer Berührungen sowie fließende, menschenähnliche Bewegungen (Bartneck, Rosalia, Menges & Deckers, 2005; Ferrari & Eyssel, 2016; Marquardt, 2017, S. 30; Remmers, 2020; Złotowski, Strasser & Bartneck, 2014), die diese Cobots häufig zu Vermenschlichungsobjekten machen (vgl. Kapitel 3.2.4.1). Beim Cobot *Baxter* von *Rethink Robotics* kommt ein auffälliges Merkmal der äußeren Gestaltung hinzu. Dieser Cobot verfügt über ein Display mit einem simulierten Augenpaar, das über die simulierte Blickrichtung eine intuitive Kommunikationsmöglichkeit bietet (Fitzpatrick, Harada, Kemp & Matsumoto, 2016, S. 1790; Sheridan, 2016, S. 526). Diese Form der Kommunikation ist bei Teams aus Menschen genauso wie bei hybriden Teams aus Menschen und Robotern von großer Relevanz (Breazeal, Kidd, Thomaz, Hoffman & Berlin, 2005; Mutlu et al., 2016, S. 1913). Letztlich machen sich die Hersteller damit die menschliche Neigung zunutze, mögliche soziale Interaktionspartner:innen genau zu beobachten (Campa, 2016, S. 110) und auf soziale Hinweisreize besonders zu reagieren (S. Lee et al., 2005). So dient bspw. im zwischenmenschlichen Kontakt die Mimik intuitiv zur sozialen Interaktion und zur Zuschreibung von mentalen Zuständen (Breazeal et al., 2005, S. 713). Insbesondere die Pupillenbewegungen im Auge, die aufgrund des hohen Kontrasts zwischen Pupille und umgebender Sklera leicht zu erkennen sind, vereinfachen die Antizipation künftiger Bewegungsrichtungen fremder Personen und tragen damit zu einer gelingenden Interaktion bei (Compagna et al., 2016, S. 5). Solche intuitiven Interaktionsmechanismen sind durch das simulierte Augenpaar des *Baxter* ebenfalls angelegt, wenngleich sie noch nicht praktisch realisiert sind. Insofern handelt es sich beim Cobot *Baxter* um ein besonders prototypisches Exemplar eines hybriden Roboters, das bereits in der äußerlichen Gestaltung Mehrdeutigkeiten bezüglich der Zuordnung aufweist. Gleichwohl erfüllen u. U. auch weniger offensichtlich anthropomorph gestaltete und weit verbreitete Cobots, wie bspw. aus der UR-Reihe von *Universal Robots* (*UR*), die Kriterien für hybride Roboter.

2.3 Industrieller Arbeitsplatzkontext

2.3.1 Auswirkungen auf die Arbeitswelt

Im Zusammenhang mit Robotern im Allgemeinen und mit Cobots im Speziellen werden oftmals die Auswirkungen auf die Verfügbarkeit von Arbeitsplätzen sowie auf die Arbeitsbedingungen der Menschen thematisiert und reflektiert. In der öffentlichen Debatte stehen

sich in diesem Zusammenhang zumeist zwei stark polarisierte Zukunftsentwürfe gegenüber, die sich auch technikhistorisch verorten und rechtfertigen lassen. Bereits im 19. Jahrhundert stellten Maschinen häufig eine Konkurrenz für Mitarbeitende dar, wohingegen sie aus Sicht der Arbeitgeber:innen als Mittel zur Effizienzsteigerung und zur Entlastung wahrgenommen wurden. Mitunter begrüßten Mitarbeitende die Einführung von Maschinen aber auch, sofern sie diese als Unterstützung wahrnahmen. Bereits bei diesen frühen Technisierungsprozessen zeigten sich entsprechende „Vieldeutigkeiten“ und eine „anthropozentrische[] Angst vor Ersetzung“ (Heßler, 2019, S. 52). Während die zurückliegende Technologisierung der Arbeitswelt gemeinhin als Fortschritt bewertet wird, wird eine zukünftige Automatisierung manueller Tätigkeiten individuell als Gefahr und/oder als Chance betrachtet (Turja & Oksanen, 2019). Auf Basis der Eindrücke aus medial vermittelten Debatten oder Erzählungen von Kolleg:innen stehen Mitarbeitende der Einführung von Robotern als potenzielle Substitutionstechnologie häufig ambivalent gegenüber (Meißner et al., 2020).

Aufgrund des breiten Einsatzspektrums von Cobots jenseits konventioneller Automatisierung (Steil & Maier, 2017b) werden diese in positiven Zukunftsentwürfen bisweilen als „Ausweg aus der weiteren Vernichtung von Arbeitsplätzen durch die fortschreitende industrielle Entwicklung“ (Onnasch et al., 2019, S. 24) bezeichnet. Dadurch „adressieren MRK-Szenarien traditionelle Automatisierungsängste“ (Remmers, 2020, S. 55), was in der öffentlichen Debatte als soziale Implikation des Cobot-Einsatzes stark hervorgehoben wird. Begrifflichkeiten wie *Kollege Roboter* oder *hybride Teams* (Müller et al., 2017; Onnasch & Roesler, 2019; Richert et al., 2016) verdeutlichen, dass Cobots und deren menschliche Interaktionspartner:innen stets als gemeinsames sozio-technisches Arbeitssystem anzusehen sind (Decker, 2017, S. 154). Aus arbeitswissenschaftlicher Sicht hängt dessen Performance nicht nur von den einzelnen Fähigkeiten des beteiligten Menschen und des Roboters, sondern auch maßgeblich von deren Interaktion ab (DIN EN ISO 6385). Entsprechend verschiebt sich der Fokus in der öffentlichen Debatte von einer Substitution der Mitarbeitenden hin zu deren verstärkter Zusammenarbeit mit den jeweiligen Maschinen (Brandstetter, 2017, 17f.; Decker et al., 2017, S. 348; Oliveira et al., 2018). Gleichwohl verläuft in der Praxis die Grenze zwischen einem assistierenden, einem kompetitiven und einem substituierenden Roboter unscharf, da sich eine unterstützende Technologie im Laufe der Zeit zu einer ersetzenden entwickeln kann (Borenstein, 2011, S. 87).

Neben der diffizilen Unterscheidung zwischen Unterstützungs- und Ersetzungstechnologie wird die Debatte um die Rolle von Robotern und deren Relation zu Mitarbeitenden im betrieblichen Kontext traditionell anhand des Dualismus zwischen Werkzeug und Teamkollege (*tool vs. teammate*) geführt. Dabei wird überhaupt erst seit der Entwicklung von Robotern ernsthaft in Betracht gezogen, dass Mitarbeitende eine Technologie nicht als bloßes Werkzeug, sondern als Teammitglied wahrnehmen könnten (Groom & Nass, 2007, S. 486). Groom und Nass (2007) definierten bereits 2007 sechs Kriterien für eine:n

„erfolgreiche:n Teamkolleg:in“ (*successful teammate*; S. 483ff.). Dazu zählten sie u. a. die Fähigkeit, geteilte mentale Modelle zu entwickeln und sich gegenseitig zu vertrauen. Da Roboter die mentalen Modelle ihres menschlichen Gegenübers nicht ergründen und demnach keine geteilten mentalen Modelle entwickeln können, schlussfolgerten die Autor:innen, dass Roboter nicht als adäquate Teammitglieder erhalten (*robots fail as teammates*) (Groom & Nass, 2007, S. 489). Diese theorie- und logikbasierte Argumentation ist allerdings nicht empirisch belegt. Eine der ersten empirischen Veröffentlichungen zur Wahrnehmung von Robotern als Kooperationspartner:innen bei der Bewältigung von Aufgaben stammt von Hinds et al. (2004). Die Autor:innen fanden hierbei in einem Laborexperiment Hinweise darauf, dass Versuchspersonen mehr Verantwortung bei der Durchführung einer Aufgabe abgeben, wenn sie diese gemeinsam mit einem menschenähnlichen anstatt mit einem maschinenähnlichen Roboter erledigen.

Insbesondere die historischen Betrachtungen sowie die Analyse der öffentlichen Debatten beziehen sich aufgrund der Neuartigkeit von Cobots nicht notwendigerweise konkret auf diesen Robotertyp, sondern auf generellere Mensch-Maschine- und Mensch-Roboter-Relationen. Da Cobots ihrer Bezeichnung nach für den kollaborierenden Einsatz konstruiert und intendiert sind, ließe sich zunächst vermuten, dass sie weniger Ängste vor technologischer Ersetzung auslösen (Sauppé & Mutlu, 2015, S. 3613). Da in realen Szenarien Cobots allerdings häufig als Leichtbauroboter für koexistente oder kooperative Anwendungszenarien angesehen werden (vgl. hierzu bspw. den Anwendungsfall in El Makrini et al., 2018), erscheint eine gewisse Sorge vor einem Arbeitsplatzverlust durchaus berechtigt. Letztlich wird sich im praktischen Einsatz vor dem Hintergrund behördlicher Regulierungen und gesellschaftlicher sowie firmeninterner Akzeptanz entscheiden, ob Cobots entweder zur Vollautomatisierung von Teilprozessen oder zur Unterstützung der Mitarbeitenden und zur ergonomischeren Gestaltung von Arbeitsplätzen führen werden (Steil & Maier, 2017b, S. 407).

2.3.2 Betriebliche Cobot-Einführungen

2.3.2.1 Phasen des Einführungsprozesses

Betriebliche Cobot-Einführungsprozesse stellen für Unternehmen eine komplexe Herausforderung dar. Zunächst bedingt die zuvor geschilderte Konzeption von Mitarbeitenden und Cobots als sozio-technisches System mit einer Vielzahl an Wechselwirkungen, dass auch der Einführungsprozess eines Cobots eine sozio-technische Aktivität darstellt, die den Einbezug der Mitarbeitenden erfordert (Erickson, Robert, Crowston & Nickerson, 2018, S. 359; Frennert, 2019; Moniz & Krings, 2014, S. 288). Dabei sind Erfahrungen und Erkenntnisse aus ähnlichen Technologieadoptionsprozessen und Mensch-Maschine-Interaktionen nicht ohne Weiteres auf diese Prozesse übertragbar. Vielmehr ist es geboten, stets die beteiligten Menschen und Maschinen individuell zu betrachten, da sich das

menschliche Selbstbild in der Konfrontation mit verschiedenen Maschinentypen unterschiedlich verändert. Die Verwendung von *Mensch* und *Maschine* im Kollektivsingular unterschlägt indes die Heterogenität verschiedener Menschen und Maschinen (Heßler, 2019). Gerade MRI verändert die „Praktiken, Selbstverständnisse und die Relation von Menschen und Maschinen“ (Heßler, 2019, S. 46f.) in neuer Art und Weise und in Abhängigkeit davon, welche Menschen und Roboter konkret beteiligt sind.

In den vergangenen Jahren sind verschiedene Rahmenwerke, morphologische Ansätze und Referenzmodelle entstanden, um den Einführungsprozess von Cobots im betrieblichen Umfeld abzubilden (Görke et al., 2017; Malik & Bilberg, 2017, 2019; Ranz et al., 2018). Nur wenige Modelle legen dabei eine chronologische Betrachtung entlang verschiedener Phasen des Einführungsprozesses zugrunde. Ein Beispiel dafür stellt das Modell nach Malik und Bilberg (2017) dar, das zwischen einer Konzeptentwicklungs-, einer Explorations- und einer Entscheidungsphase differenziert. Dieses Modell bezieht sich demnach auf die Entscheidungsfindung für oder gegen einen Cobot, auf die Auswahl eines geeigneten Produkts und auf die Entwicklung eines Konzepts für die Implementierung an einem bestehenden Arbeitsplatz. Die tatsächliche Implementierung sowie die Betriebsphase des Cobots bleiben bei dieser Betrachtung außen vor.

Einen weiteren relevanten Blickwinkel bietet der chronologisch strukturierte und empirisch fundierte Ansatz von Wurhofer, Meneweger, Fuchsberger und Tscheligi (2015), der auf den Ergebnissen einer qualitativen *in situ*-Studie mit zehn Mitarbeitenden eines produzierenden Unternehmens basiert. Die Forschenden differenzieren den Einführungsprozess eines Cobots aus Sicht der beteiligten Mitarbeitenden in drei Phasen, die sich am Umfang der Interaktionserfahrungen mit dem Cobot orientieren. Diese Phasen lassen sich typischen unternehmerischen Phasen bei der Technologieeinführung zuordnen, nämlich erstens einer Planungsphase, zweitens einer Implementierungsphase und drittens einer Betriebsphase. In der Planungsphase werden geeignete Anwendungsfälle identifiziert und Umsetzungs-ideen ausgearbeitet. Die Mitarbeitenden haben zu diesem Zeitpunkt noch nicht mit dem betreffenden Cobot interagiert. Die Implementierungsphase umfasst die Installation und Konfiguration des Cobots und dessen benötigter Peripherie am konkreten Arbeitsplatz und ermöglicht den Mitarbeitenden erste Interaktionserfahrungen. Die Betriebsphase beschreibt den Zeitraum, in dem der installierte Cobot in den laufenden Betrieb eingegliedert ist und alltägliche Nutzungserfahrungen und damit einen ersten Abgleich der Erwartungen mit tatsächlichen Erfahrungen ermöglicht. Die drei Phasen mit den entsprechenden Erkenntnissen aus der Fallstudie sind in Abbildung 2.5 visualisiert. Dabei ist zu beachten, dass der Prozess der Technologieeinführung in der Praxis nicht idealtypisch sequenziell abläuft, sondern Rückkopplungsschleifen und Iterationen beinhaltet, die aus Gründen der Komplexitätsreduktion in der schematischen Abbildung nicht dargestellt sind.

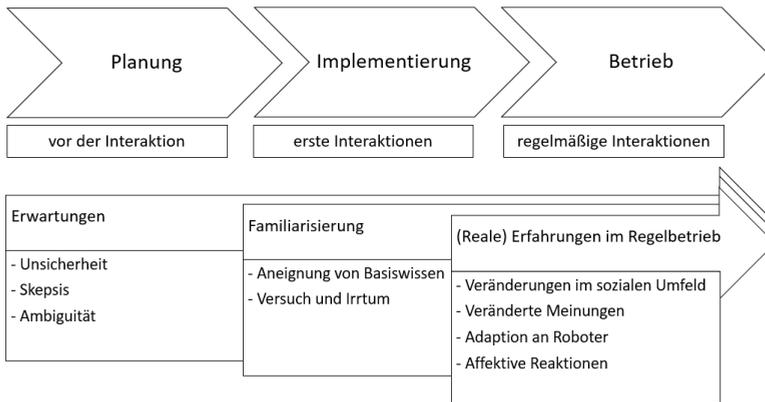


Abbildung 2.5: Phasen der Cobot-Einführung aus Sicht der betroffenen Mitarbeitenden im zeitlichen Veränderungsverlauf(stark angelehnt an Wurhofer et al., 2015).

In der frühen Phase vor dem ersten Kontakt mit einem Cobot beschäftigten sich die Mitarbeitenden laut der Studie von Wurhofer et al. (2015) mit ihren Erwartungen in Hinblick auf die spätere Interaktion und auf die damit einhergehenden Veränderungen an ihrem Arbeitsplatz. Bereits zu diesem Zeitpunkt fühlten sich die befragten Mitarbeitenden von den anstehenden Veränderungen betroffen, versuchten sich zu informieren, beteiligten sich an Gerüchten und begannen, sich eine Meinung zu bilden. Diese Phase war gekennzeichnet von Unsicherheit, einem hohen Maß an Skepsis gegenüber der Einführung sowie ambivalenten Erwartungen. Dabei spielte es auch eine Rolle, wie sich die übrigen Mitarbeitenden zur Cobot-Einführung positionierten, die bereits erste Interaktionserfahrungen gesammelt hatten. Die folgende Phase der Familiarisierung beinhaltete die Reorganisation der persönlichen Arbeitsvorgänge und räumlichen Anordnungen am Arbeitsplatz. Es entstanden erste Erfahrungen in der Interaktion mit dem Cobot; auf Basis von Versuch und Irrtum bildete sich eine Wissensbasis heraus. In der dritten Phase zeigten sich reale Konsequenzen der Einführung. Hierzu zählten auch Veränderungen im sozialen Umfeld am Arbeitsplatz, da Cobot-Techniker:innen häufig zu neuen Kolleg:innen wurden und sich vermehrt Mitarbeitende aus anderen Bereichen bei den Mitarbeitenden mit direktem Kontakt zum Cobot nach deren Erfahrungen erkundigten. Skeptische Haltungen verflüchtigten sich, sodass der Gesamteindruck von der Cobot-Einführung positiver wurde, zumal sich die initialen Befürchtungen bspw. bezogen auf den Verlust des eigenen Arbeitsplatzes im konkreten Fall nicht bewahrheiteten. Insgesamt zeigten sich in der Studie vor allem in der frühen, von subtilen Emotionen und Erwartungen geprägten Phase besonders kritische Meinungen gegenüber Cobots (Wurhofer et al., 2015). Gezielte Trainingsmaßnahmen können nach Meinung der Autor:innen positive erste Nutzungserfahrungen begünstigen, die zu einem Gefühl der Kompetenz führen und potenziell weitere Zuschreibungen prägen (Wurhofer et al., 2015,

S. 216). Einschränkend ist zu erwähnen, dass das Modell nur auf einer Einzelfallstudie basiert. Nach dieser zeitlichen Einordnung werden im folgenden Kapitel die Wahrnehmungen und Reaktionen von Mitarbeitenden auf Cobot-Einführungen anhand weiterer Studien detaillierter beschrieben.

2.3.2.2 Wahrnehmungen und Reaktionen der Mitarbeitenden

Es existieren nur wenige praxisnahe Studien, die sich damit beschäftigen, wie Mitarbeitende die Einführung von Cobots im industriellen Kontext wahrnehmen und welche psychischen Konsequenzen sich daraus ergeben (Steil & Maier, 2017b, S. 404; Weiss & Huber, 2016). Eine der wenigen Studien, die auf Mitarbeitende in produzierenden Unternehmen als reale Endanwender:innen von Cobots zurückgreift, stammt von Elprama et al. (2017). Mittels qualitativer Expert:inneninterviews befragten die Forschenden acht Fabrikmitarbeitende, die zuvor mit dem Cobot *Baxter* eine Montageaufgabe kollaborierend bearbeitet hatten. Die Produktionsmitarbeitenden vermuteten, dass der Cobot ihre mentale und physische Beanspruchung senke, allerdings auch den zwischenmenschlichen Kontakt reduziere. Trotz der physischen Nähe zum Cobot fühlten sich die Versuchspersonen sicher im Umgang mit *Baxter*. Ferner identifizierten sie einige vorteilhafte Aspekte des Cobots, z. B. die ergonomische Übergabe von Werkstücken an Mitarbeitende in deren bevorzugter Greifhöhe. Dennoch stellten die Mitarbeitenden in Frage, ob sich mithilfe des Cobots die Arbeitsaufgaben wirklich schneller erledigen ließen. Außerdem äußerten sie Bedenken, langfristig durch *Baxter* ersetzt zu werden und ihren Arbeitsplatz zu verlieren (Elprama et al., 2017).

Eine weitere Untersuchung des Einsatzes von Cobots im Produktionsbetrieb liefern Sauppé und Mutlu (2015). Mittels non-reaktiver Beobachtung und semi-strukturierten Interviews mit Vertreter:innen verschiedener Unternehmensebenen (Operator:innen, Instandhalter:innen, Management) untersuchten sie zwei KMU sowie ein internationales Großunternehmen, die ebenfalls den Cobot *Baxter* eingeführt hatten. Die Autor:innen schlussfolgern auf Basis ihrer Beobachtungen, dass die Mitarbeitenden die Cobots menschenähnlich repräsentieren, ihnen eine Persönlichkeit und Absichten zuschreiben, sozial mit ihnen interagieren und sich aufgrund des menschenähnlichen Erscheinungsbildes in näherer Umgebung des Roboters sicher fühlen. Vermenschlichungsprozesse (vgl. Kapitel 3.2) scheinen in Bezug auf hybride Roboter in der Produktionsumgebung demnach stattzufinden (Sauppé & Mutlu, 2015) und mitunter zur Akzeptanz innerhalb der Belegschaft beizutragen (Oistad et al., 2016, S. 783). Die Vermenschlichung kann sich dabei darin äußern, dass den Cobots menschliche Namen verliehen werden (Šabanović, Reeder & Kechavarzi, 2014).

In einer Befragung von El Makrini et al. (2018) unter 42 Mitarbeitenden in einer Automobilproduktion spiegelte sich erneut die Angst vor Arbeitsplatzverlust wider. Eine Reduktion der Arbeitsplätze wurde hierbei als größter Nachteil des Robotereinsatzes genannt (23 %). Zusätzlich bezeichneten 10 % der Befragten die Ersetzung der Mitarbeitenden

durch Roboter als Nachteil. Bei dieser Sorge handelt es sich um ein wesentliches Akzeptanzhemmnis, das bereits seit der ersten industriellen Revolution gegenwärtig ist und im Zuge neuer Technologie neue Facetten gewinnt (Salvini et al., 2010, S. 456). Das zeigt sich auch in einer Studie von Maurtua, Iburguren, Kildal, Susperregi und Sierra (2017), in der die 115 Versuchspersonen mit Erfahrung im Bereich der Industrieroboter vielerlei Vorzüge von Cobots anerkannten, aber mehrheitlich der Meinung waren, dass Cobots einen negativen Einfluss auf die Anzahl der Jobs haben werden. Solche Ängste vor dem Verlust des eigenen Arbeitsplatzes durch Robotereinführungen lassen sich auch in größer angelegten repräsentativen Befragungen messen, wenngleich in Kombination mit durchaus positiven Erwartungshaltungen. Im sog. *Eurobarometer* (European Commission, 2015) betrachteten 70 % der Befragten Roboter als Jobkiller, gleichzeitig bewerteten allerdings 85 % der Studienteilnehmenden diese als notwendige Hilfe bei schweren Arbeiten und 72 % als gute Sache für die Gesellschaft. Das lässt sich als Indiz dafür ansehen, dass zwischen beiden Wahrnehmungen nur ein schmaler Grat verläuft (Weiss et al., 2016). Dabei bleibt häufig unklar, wie sich die Mitarbeitenden zum Cobot in Beziehung setzen, d. h. ob sie ihn jeweils als unterlegen, gleichwertig oder überlegen erachten bzw. als unterstützend oder bedrohend erleben (Oliveira et al., 2018, S. 280). Wenngleich diese relationale Verortung auf falschen Annahmen beruhen und aus der neutralen Außenperspektive als unzutreffend erscheinen kann, wirkt sie sich wesentlich auf das Vertrauen in den Cobot und auf die Akzeptanz desselben aus (Meißner & Trübswetter, 2018, S. 227; Meißner et al., 2020). Insbesondere, da Technologieakzeptanz und -vertrauen von Mitarbeitenden nicht ausschließlich aus rationalen oder analytisch geprägten Prozessen resultieren, sind praxisnahe empirische Studien dem umfassenden Verständnis dieser Konstrukte besonders dienlich (Meißner et al., 2020).

Während über die arbeitsmarktbezogenen und volkswirtschaftlichen Folgen der Technisierung der Arbeit einige Studien vorliegen², sind die unmittelbaren psychischen Folgen des Arbeitsplatzverlusts auf die jeweilig betroffenen Mitarbeitenden bisher kaum untersucht (Granulo et al., 2019). Um diese Lücke zu schließen, streben Granulo et al. (2019) die Begründung einer Psychologie der technologischen Ersetzung (*psychology of technological replacement*) an. Sie stützen sich dabei auf Erkenntnisse aus einer aufwändigen Serie von elf Experimenten mit unterschiedlichen Stichproben, die neben Studierenden auch Produktionsmitarbeitende umfassten. In diesen Experimenten untersuchten die Autoren, wie die Ersetzung durch eine neue Technologie im Vergleich zur Ersetzung durch neue Mitarbeitende bewertet bzw. mental verarbeitet wird. Dabei zeigte sich, dass die Bewertung von der eingenommenen Perspektive (*perspective taking*) abhängt, nämlich davon, ob die Person entweder selbst mit einer Ersetzung konfrontiert ist (Betroffenen-Perspektive) oder ob sie die Ersetzung einer Kollegin bzw. eines Kollegen beobachtet (Beobachtenden-

² Besondere mediale Aufmerksamkeit erregte dabei in den letzten Jahren bspw. die viel rezipierte US-amerikanische Arbeitsmarktstudie von Frey und Osborne (2013) und die zugehörige Übertragungsstudie für Deutschland von Bonin, Gregory und Zierahn (2015).

Perspektive). In der Beobachtenden-Perspektive bevorzugen es Mitarbeitende, dass betroffene Kolleg:innen durch neue menschliche Mitarbeitende ersetzt werden. Das begründen Granulo et al. (2019) mit prosozialen Gefühlen. In der Betroffenen-Perspektive, bei der der eigene Arbeitsplatz bedroht ist, ziehen es Arbeitnehmer:innen vor, durch eine Technologie ersetzt zu werden. Für diese Präferenz machen die Autoren soziale Vergleichsprozesse gegenüber Mitmenschen verantwortlich. Der Eindruck, schlechter als ein Mitmensch geeignet zu sein, bedroht und beschädigt dabei mitunter den eigenen Selbstwert und kann zu einer negativen Bewertung des überlegenen Konkurrenten führen. Die Fähigkeiten einer Maschine werden hingegen als inkommensurabel angesehen. Da Technologien gezielt dazu entwickelt werden, Menschen in einem bestimmten Anwendungsbereich zu übertreffen (Grunwald, 2019a, S. 33), wird die Unterlegenheit gegenüber der Maschine nicht als persönlicher Makel aufgefasst, sondern verweist auf eine grundlegende Leistungsdifferenz zwischen Menschen und Maschinen. Besonders bemerkenswert ist dabei, dass im Falle einer starken Vermenschlichung einer Maschine die gleichen Vergleichsprozesse abzulaufen scheinen wie gegenüber Mitmenschen (Yogeeswaran et al., 2016). Erst in einer langfristigen Betrachtungsperspektive scheint die Ersetzung durch eine nicht-vermenschlichte Technologie nachteilig, da sie mit einer stärkeren ökonomischen Bedrohung, also einer geringen Wahrscheinlichkeit, eine vergleichbare Anstellung in einem anderen Unternehmen zu finden, assoziiert wird. Davon unbenommen bleibt, dass aus einer kurzfristigen Perspektive die Ersetzung durch eine Technologie präferiert wird (Granulo et al., 2019).

Die bisherigen Betrachtungen fokussierten die Ebene eines einzelnen Arbeitsplatzes und die Reaktionen der an diesem Arbeitsplatz beschäftigten Mitarbeitenden. Studien aus dem Bereich der Service-Robotik verdeutlichen allerdings, dass sich die Einführung eines Roboters sogar über den konkret betroffenen Arbeitsplatz hinaus auswirken und in einem schleichenden Prozess die Organisationskultur verändern kann (Jung & Hinds, 2018; M. K. Lee, Kiesler, Forlizzi & Rybski, 2012). Charles et al. (2015) vertreten die Auffassung, dass die Einführung eines Cobots komplexe sozialpsychologische Phänomene zur Folge haben kann. Dazu müssen die jeweiligen Roboter nicht zwangsläufig anthropomorph gestaltet sein. Vielmehr zeigen Mitarbeitende auch anderen Robotern gegenüber soziale Interaktionsmuster, sofern sie diese bspw. aufgrund persönlicher Überzeugungen oder medial vermittelter Vorstellungen als sozial repräsentieren (Stenzel, Chinellato, Bou et al., 2012). Umgekehrt bietet die Organisationsstruktur ein unterschiedlich gut geeignetes Umfeld für solche Einführungsprozesse (Shehadeh et al., 2017), sodass die Cobot-Einführung als sozio-technischer Prozess auch bidirektionale Wechselwirkungen mit der Organisationsstruktur einbeziehen sollte.

Abschließend ist relativierend hinzuzufügen, dass es sich bei den vorliegenden Untersuchungsergebnissen zur Reaktion von Mitarbeitenden auf Cobot-Einführungen in aller Regel nur um erste Indizien handelt, die noch unzureichend empirisch abgesichert sind. Das liegt daran, dass die Studien aufgrund der schwierigen Adressierbarkeit der Zielgruppe und

der begrenzten Verbreitung von Cobots entweder in Laborsituationen stattfinden, die erheblich von den realen Anwendungskontexten abweichen (mangelnde externe Validität), oder – falls sie in realen Umgebungen durchgeführt werden – häufig Fallstudien-Charakter mit sehr geringen Stichprobengrößen aufweisen (mangelnde Repräsentativität).

2.4 Gegenstandsbereich dieser Dissertation

Nach den vorgenannten Unterscheidungen lässt sich nun der Gegenstandsbereich der vorliegenden Dissertation exakter bestimmen. Dabei ist zu beachten, dass die Unterscheidungsdimensionen weitestgehend unabhängig voneinander, einzelne Unterscheidungen allerdings nicht immer trennscharf sind.

Forschungsdisziplin:	Kognitive MRI	Physische MRI
Mensch-Cobot-Arbeitssystem:	Mitarbeitende	Interaktion Cobot
Anwendungskontext:	Industrielle Robotik	Service-Robotik
Robotertyp:	Kollaborationsfähige Roboter	Klassische Industrieroboter
Interaktionstyp:	Kollaboration	Kooperation Koexistenz
Robotergestaltung:	Funktionale Roboter	Hybride Roboter Soziale Roboter
Cobot-Einführungsprozess:	Phase der Erwartungen (vor der Interaktion)	Phase der Familiarisierung (erste Interaktionen) Phase der Erfahrungen (regelm. Interaktionen)

Abbildung 2.6: Bestimmung des Gegenstandsbereichs dieser Dissertation anhand vorgenannter Unterscheidungsdimensionen.

Wie Abbildung 2.6 zu entnehmen ist, analysiert die vorliegende Dissertation Interaktionen aus einem primär kognitionswissenschaftlichen Blickwinkel im Rahmen der kognitiven MRI (Forschungsdisziplin) und fokussiert hierbei die mentalen Modelle der Mitarbeitenden als Bestandteil des Mensch-Cobot-Arbeitssystems. Dabei konzentriert sich die Dissertation auf im industriellen Umfeld (Anwendungskontext) eingesetzte kollaborationsfähige Roboter (Robotertyp). Diese industriellen Cobots erfüllen eine Aufgabe innerhalb des Produktionsprozesses eines Unternehmens und sind daher keine ausschließlich sozialen Roboter. Dennoch können diese Cobots aufgrund ihres hohen Maßes an Interaktivität sowie ihrer mitunter vorhandenen sozialen Funktionsmerkmale als hybride Roboter begriffen werden (Robotergestaltung). Im Zentrum der Forschung steht deren Anwendung in kollaborierenden oder kooperativen Einsatzszenarien (Interaktionstyp). Aufgrund der geringen Interaktivität in koexistenten Szenarien bleiben diese unberücksichtigt, da bei dieser Interaktionsform nur geringe Auswirkungen auf die menschenzentrierten Faktoren zu erwarten

sind. In Hinblick auf die zeitliche Dimension stehen die frühen Phasen des Einführungsprozesses im Vordergrund, in denen die Mitarbeitenden erste Erwartungen an den Cobot ausbilden und sich in ersten Interaktionen mit dem Cobot vertraut machen (Cobot-Einführungsprozess).

Nach diesen Begriffsbestimmungen zur industriellen MRI sowie der Eingrenzung des Gegenstandsbereichs dieser Dissertation werden im nächsten Kapitel zentrale theoretische Konzepte aus dem Bereich der kognitiven MRI erläutert, in Zusammenhang gebracht und auf die zentralen Forschungsfragen hin verdichtet.

3 Kognitive Mensch-Roboter-Interaktion

Dieses Kapitel widmet sich vier zentralen Themenkomplexen, die der kognitiven MRI zuzuordnen sind. Kapitel 3.1 charakterisiert mentale Modelle als komplexitätsreduzierende interne Repräsentationen, die u. a. die Vorstellung von Cobots prägen. Diese gedanklichen Strukturen sind relevant, um das in Kapitel 3.2 thematisierte Phänomen des Anthropomorphismus, also die Neigung zur Vermenschlichung technischer Artefakte, zu erklären. In Kapitel 3.3 wird anschließend sprachliches Framing als Mechanismus beschrieben, der mittels sprachlicher Variationen imstande ist, bestimmte mentale Modelle zu aktivieren und damit Vermenschlichungsprozesse zu initiieren. Diese wirken sich wiederum auf das Mensch-Roboter-Vertrauen aus, das in Kapitel 3.4 erläutert wird. Innerhalb der einzelnen Kapitel werden die jeweiligen Themen zunächst allgemein beschrieben und anschließend auf den MRI-Kontext bezogen. Zuletzt erfolgt in Kapitel 3.5 eine Zusammenführung der Themenfelder, die in die Ausformulierung von sechs Forschungsfragen und in die Spezifikation des empirischen Forschungsprogramms mündet.

3.1 Mentale Modelle

„Der Mensch ist des Denkens nicht in hohem Maße fähig, und auch noch der geistigste und gebildetste Mensch sieht die Welt und sich selbst beständig durch die Brille sehr naiver, vereinfachender und umlügender Formeln an – am meisten aber sich selbst!“
– Hesse, 1927/2016, S. 76 –

3.1.1 Begriffsbestimmung

Diese Brille, die Herman Hesse in seinem Roman *Der Steppenwolf* anspricht und die einen vereinfachenden Blick auf die Welt ermöglicht, deutet bereits die mentalen Modelle an, die dem menschlichen Denken als komplexitätsreduzierende Heuristik zugrunde liegen. Das Konzept des mentalen Modells geht auf Kenneth Craik zurück, der selbiges 1943 in die psychologische Forschung einbrachte (Craik, 1967/1943, S. 61). Er formulierte die Hypothese, dass Menschen wahrgenommene äußere Objekte und Zusammenhänge in komplexitätsreduzierenden mentalen Modellen (*small-scale models*) abbilden, um diese intern zu manipulieren, verschiedene Szenarien zu simulieren und Zukunftsvorhersagen abzuleiten (Craik, 1967/1943, S. 61; Johnson-Laird, 2005, S. 186; Johnson-Laird & Khemlani,

2017, S. 170). Nach der kognitiven Wende, die den Behaviorismus als vorherrschende Strömung der Psychologie durch den Kognitivismus ablöste, rückte das Konzept der mentalen Modelle in den 1980er-Jahren verstärkt in den Fokus der (empirischen) Forschung (Johnson-Laird & Khemlani, 2017, S. 170; Seel, 2017).

Heutzutage besteht in den Kognitionswissenschaften und der Psychologie weitgehend ein Konsens darüber, dass Menschen solche mentalen Modelle entwickeln und mit ihnen operieren (Held, Knauff & Vosgerau, 2006, S. 5f.; Jones et al., 2011). Die Anwendungsfelder sind dabei vielfältig. So gilt die Theorie der mentalen Modelle als derzeit beste Theorie zur Erklärung des räumlichen Denkens (Ragni & Knauff, 2013), findet aber auch im Bereich des Schlussfolgerns (mentale Logik; Johnson-Laird, 2010), des Sprach- und Textverstehens (Hemforth & Konieczny, 2006), der Bildungswissenschaft und der Pädagogik (Ifenthaler & Seel, 2012; Seel, Ifenthaler & Pirnay-Dummer, 2009) Anwendung. Mentale Modelle werden dabei als mentale visuoräumliche Gebilde sowie als Repräsentanzen von Text- oder Lerninhalten betrachtet, die im Rahmen des Lernprozesses Modifikationen erfahren (*model-based learning*; Seel, 2017). Mitunter werden auch sog. kognitive Karten als Spezialform räumlicher mentaler Modelle betrachtet (Jones et al., 2011).

3.1.2 Merkmale und Funktionen

Zuvorderst handelt es sich bei mentalen Modellen um eine bestimmte Art von Modellen. Damit ist allerdings noch keine hinreichend konkrete Bestimmung gegeben, denn der Modell-Begriff gilt als „Wort-Joker“ (Janich, 2002, S. 15). Modelle lassen sich grundsätzlich aus zwei Perspektiven betrachten. Einerseits besitzen sie im Sinne des *Modells von* etwas eine deskriptive und andererseits im Sinne des *Modells für* etwas eine präskriptive Funktion, insofern sie als Heuristik für die Antizipation künftiger Geschehnisse herangezogen werden und eigene Handlungen anleiten (Gutmann & Nick, 2019, S. 209). Ein *Modell von* etwas verfolgt das Ziel einer möglichst akkuraten Abbildung eines Betrachtungsgegenstandes (*modellandum*) und basiert folglich auf der Prämisse, dass eine von den Beobachtenden unabhängige und sinnlich erfassbare Realität existiert. Dieses Paradigma verfügt über eine lange philosophische Tradition (Janich, 2002, S. 18ff.). Aus konstruktivistischer Sicht bilden mentale Modelle hingegen keine extern verfügbaren Strukturen ab, sondern nehmen selbst eine konstruierende Strukturierung der externen Welt vor. Einerseits entstehen sie aus einem Bottom-Up-Verarbeitungsprozess der Umweltwahrnehmung. Andererseits strukturieren und prägen sie diese Wahrnehmung gleichzeitig in einem Top-Down-Prozess und bilden damit eine Art Gleichgewicht zwischen innerer und äußerer Welt (Seel, 2001, S. 408). Dabei können neben (physischen) Objekten insbesondere auch abstrakte, visuell nicht erfassbare Phänomene, Konzepte, Strukturen und raumzeitliche Zusammenhänge zum Gegenstand dieser Modelle werden (Johnson-Laird, 2005, 2010; Johnson-Laird & Khemlani, 2017, S. 177). Gleichwohl handelt es sich bei mentalen Modellen im Gegensatz

zu propositionalen Strukturen um analoge, ikonische Repräsentationen der äußeren Welt, die deren Struktur möglichst gleichartig abbilden (Johnson-Laird, 2005, 2010). Historisch finden sich hierfür Vorläufer bspw. in der sog. *Theorie des Bildlichen* nach Wittgenstein, die den Begriff des *Abbildes* nicht rein visuell versteht (Held et al., 2006, S. 13; Johnson-Laird, 2005, S. 186).

Ein perfektes *Modell von* etwas würde allerdings lediglich eine mentale Verdopplung des *modellandum* darstellen und somit nicht über eine komplexitätsreduzierende Wirkung verfügen. Für die heuristische Übertragung von Eigenschaften und Wirkzusammenhängen auf andere Objekte und für die Antizipation von zukünftigen Zuständen wäre es kaum geeignet. Dieser Zweck wird mit dem Begriff des *Modells für* etwas beschreiben. Während das *Modell von* stets einen deutlichen Bezug zu einem empirisch feststellbaren Objekt aufweist, muss ein derartiges *modellandum* beim *Modell für* nicht vorliegen. Vielmehr können anhand dieser Betrachtungsweise bspw. vergangene Lebensformen oder abstrakte Zusammenhänge modelliert werden (Gutmann & Nick, 2019, S. 211ff.). Tabelle 3 bietet einen Überblick über die Funktionen, die verschiedene Autor:innen mentalen Modellen zuschreiben.

Tabelle 3: Funktionen von mentalen Modellen gemäß verschiedener Autor:innen.

Gutmann und Nick (2019)	Rouse, Cannon-Bowers und Salas (1992)	Seel et al. (2009, S. 6)
Deskriptive Funktion (<i>Modell von</i>)	Beschreibung	Mentale Repräsentation externer Gegebenheiten (statische Betrachtung)
Präskriptive Funktion (<i>Modell für</i>)	Erklärung Vorhersage	Simulation kontrafaktischer Zustände (dynamische Betrachtung)

In der Praxis lässt sich nur schwer differenzieren, ob ein mentales Modell als *Modell von* oder als *Modell für* etwas zu verstehen ist. Beide Sichtweisen schließen sich nicht aus, sondern bedingen sich wechselseitig. Die Güte eines Modells bemisst sich allerdings nach unterschiedlichen Kriterien, je nachdem, ob die deskriptive oder die präskriptive Funktion des Modells im Vordergrund steht. Beim *Modell von* etwas gilt der Grad an Übereinstimmung mit dem zu modellierenden Objekt als entscheidendes Kriterium, während beim *Modell für* etwas die Akkuratheit daraus ableitbarer Übertragungen und Vorhersagen entscheidend ist. Daher können leicht Missverständnisse entstehen, wenn Modelle primär als Erklärungsheuristik gedacht sind, aber als exakte Abbilder der Realität verstanden werden. So dienten ursprünglich elektrochemische *Modelle* als Erklärungsmuster *für* die Leistung

neuronaler Netzwerke im Gehirn. Inzwischen wird der Begriff des neuronalen Netzes in der Informatik allerdings so verwendet, dass der Eindruck entstehen könnte, als stellten diese programmierten Netze ein akkurates Abbild, also ein *Modell vom Zusammenspiel* der Neuronen im menschlichen Gehirn dar (Gutmann & Nick, 2019, S. 215f.).

Diese allgemeinen Merkmale von Modellen lassen sich auf mentale Modelle anwenden. Nach Dutke (1994, S. 4) verfügen mentale Modelle über drei zentrale Merkmale, die sich bereits aus dem allgemeinen Modellbegriff ergeben. Erstens bilden mentale Modelle äußere Gegebenheiten selektiv mental ab (Abbildungsmerkmal). Das Adjektiv *mental* weist dabei darauf hin, dass es sich um innere Strukturen handelt, die allenfalls durch Introspektion zugänglich und nicht direkt durch Dritte beobachtbar sind (Rouse & Morris, 1986; Seel, 2017). Zweitens bleiben mentale Modelle aufgrund des begrenzten Arbeitsgedächtnisses, aus Gründen der kognitiven Ökonomie und aufgrund des Charakters eines *Modells für* stets unvollständig und ungenau (Verkürzungsmerkmal). Diese Unvollständigkeit und Ungenauigkeit (*inaccuracy*) ist dabei allerdings nicht als defizitär, sondern hinsichtlich der präskriptiven Funktion des *Modells für* sogar als zweckdienlich zu betrachten. Wenngleich mentale Modelle nicht frei von Widersprüchen sind und Quellen für Fehleinschätzungen darstellen, können sie dennoch effektiv in der Anwendung sein (Ososky, 2013, S. 11; Seel et al., 2009, S. 25). Drittens werden mentale Modelle situativ für mentale Operationen wie die Manipulation des repräsentierten Objekts genutzt (pragmatisches Merkmal) und verfügen somit stets über ein „kontrafaktische[s] Moment“ (Gutmann & Nick, 2019, S. 216).

In ihrer präskriptiven Funktion bzw. aus einer dynamischen Sichtweise bilden mentale Modelle den Ausgangspunkt für mentale Simulationen von Szenarien und damit die Basis für (qualitative) Vorhersagen, Schlussfolgerungen und Entscheidungen (Denzau & North, 1994, S. 4; Johnson-Laird, 2005, S. 185; Ososky, 2013, S. 10; Seel et al., 2009, S. 6; Staggers & Norcio, 1993, S. 592). Im Stile eines Gedankenexperiments lassen sich verschiedene, kontrafaktische Umstände und Aktionen auf Basis der Modelle mental simulieren (Seel, 2017, S. 934), vergleichbar mit einer Computer-Simulation (Jones et al., 2011, S. 4), die vor dem inneren Auge abläuft (Seel, 2001, S. 407). Daher werden solche mentalen Modelle mitunter als Situationsmodelle bezeichnet (Seel, 2001, S. 407). Je mehr simulierte Zustände sich aus einem mentalen Modell ableiten lassen, desto höher fällt die kognitive Belastung aus und desto schwieriger gestaltet sich das logische Schlussfolgern (Johnson-Laird & Khemlani, 2017, S. 176). Die mentale Simulation einer Handlung kann bereits Aktionspotenziale zum Ausführen der entsprechenden Handlung provozieren (Wehling, 2016, S. 21ff.). Das veranlasste Dijksterhuis und Bargh (2001, S. 3) gar zu der These, dass die Wahrnehmung letztlich nur für das Handeln da sei (*perception is for doing*). Eindrucksvoll zeigte sich dies in einem Experiment mit Gefängnisinsass:innen, die bei einem Spiel häufiger betrogen, wenn sie vorher dezidiert zum Berichten über ihre kriminelle Vergangenheit aufgefordert wurden (Cohn, Maréchal & Noll, 2015). Das dadurch aktivierte mentale Modell, das sie als Kriminelle repräsentiert, mündete in entsprechenden Verhaltens-

skripten (vgl. auch A. Edwards, Edwards, Westerman & Spence, 2019) und provozierte kongruente kriminelle Handlungen.

Zusätzlich zu den dargelegten Funktionen und den allgemeinen Merkmalen von mentalen Modellen, die sich bereits aus dem Modellcharakter ergaben, lassen sich weitere spezifische Merkmale für mentale Modelle identifizieren. Erstens sind mentale Modelle im Sinne sog. *working models* (Craik, 1943; Johnson-Laird, 1983) stets in einem dynamischen Veränderungsprozess begriffen und daher zeitlich inkonsistent, da sie sich an neue Rahmenbedingungen anpassen und durch Lernen weiterentwickeln (Jones et al., 2011; Ososky, 2013, S. 11; Seel et al., 2009, S. 25). Zweitens sind mentale Modelle in aller Regel interindividuell stark unterschiedlich, da sich durch die Selektivität in der Abbildung, den Einfluss individueller Erfahrungen und kognitiver Verarbeitungsweisen stets Freiheitsgrade für die Ausgestaltung ergeben (Preim & Dachzelt, 2010, S. 94; Stocké, 2002, S. 47). Drittens verfügen mentale Modelle ungeachtet der grundsätzlichen Unvollständigkeit jeweils über eine gewisse interne Kohärenz, d. h. die repräsentierten Teilaspekte des Modells stehen zueinander in einem möglichst widerspruchsfreien Sinnzusammenhang (Preim & Dachzelt, 2010, S. 94).

Zusammenfassend lassen sich mentale Modelle als zwar unvollständige und vereinfachende, aber dennoch situativ gültige mentale Repräsentationen der „Verhältnisse in der Handlungsumwelt“ (Stocké, 2002, S. 46) beschreiben (*Modell von*), auf deren heuristischer Basis sich Erwartungen über die Wirkzusammenhänge und den Fortgang der aktuellen Situation angesichts des bestehenden Informationsmangels ableiten lassen (*Modell für*) (Stocké, 2002, S. 47f.). Dabei ist es durchaus möglich, dass Menschen verschiedene mentale Modelle für eine gewisse Situation besitzen, die sie kontextspezifisch aktivieren und aus denen sie jeweils unterschiedliche Erwartungshaltungen ableiten (Kiesler, 2005, S. 730f.).

3.1.3 Abgrenzungen

Die breite Anwendbarkeit mentaler Modelle zur Erklärung menschlichen Verhaltens deutet bereits auf eine weit gefasste und vage Definition hin, die die Abgrenzung zu anderen postulierten kognitiven Strukturen erschwert (Jones et al., 2011). So werden mentale Modelle bspw. mit kognitiven Schemata in Verbindung gebracht (Rouse & Morris, 1986; Wilson & Rutherford, 1989), die bereits 1932 in der Kognitionspsychologie als eine mentale Organisationsform vergangener Erfahrungen und Wissensinhalte konzeptualisiert wurden (Bartlett, 1932; Ososky, 2013, S. 10). Eine Abgrenzung von mentalen Modellen gegenüber kognitiven Schemata muss daher eher graduell denn prinzipiell ausfallen. Geeignete Kriterien nach Jones et al. (2011) sind in Tabelle 4 aufgelistet. Wie Wilson und Rutherford (1989) zusammenfassen, bestehen neben Versuchen einer Abgrenzung auch Versuche, mentale Modelle und Schemata zueinander in Beziehung zu setzen und damit als

aufeinander aufbauende Strukturen zu konzeptionieren. Eine entsprechende Annahme lautet, dass sich komplexe, situationsbezogene mentale Modelle aus verschiedenen statischen kognitiven Schemata zusammensetzen und damit die Menge aller gegenwärtig aktivierten und miteinander verknüpften Schemata repräsentieren (Wilson & Rutherford, 1989, S. 624).

Tabelle 4: Kriterien zur Abgrenzung von mentalen Modellen gegenüber kognitiven Schemata vorwiegend nach Jones et al. (2011).

Abgrenzungskriterium	Beschreibung
Dynamik	Mentale Modelle sind dynamische, im Sinne einer Simulation ausführbare Repräsentationen im Gegensatz zu statischen Datenstrukturen im Schema (Jones, Ross, Lynam & Perez, 2014, S. 1; Merrill, 2000).
Repräsentationale Flexibilität	Mentale Modelle lassen sich insbesondere auch auf neuartige Situationen anwenden, für deren Simulation verschiedene Schemata über einzelne beteiligte Entitäten verknüpft werden müssen.
Situationsbezug	Mentale Modelle sind spezifische, situationsbezogene Wissensstrukturen im Gegensatz zu weitestgehend generischen, vorkompilierten Wissensstrukturen wie Schemata.
Kognitiver Aufwand / Komplexität der Probleme	Die Anwendung mentaler Modelle zur Lösung komplexer Probleme ist ein vergleichsweise aufwändiger Vorgang, der die Konstruktion, iterative Weiterentwicklung und Simulation der Modelle einschließt, wohingegen Schemata schnell und einfach aktiviert werden. Demnach werden Schemata eher bei simplen und routinierten Problemlagen angewandt, wohingegen mentale Modelle für komplexe Probleme geeignet sind. Diese Unterscheidung ist in der Literatur allerdings umstritten (Seel, 2017, S. 942).

Häufig werden mentale Modelle außerdem mit dem visuellen Vorstellungsvermögen in Verbindung gebracht und mit mentalen Bildern gleichgesetzt. Wie bereits in Kapitel 3.1.2 angedeutet, enthalten mentale Modelle auch Informationen, die nicht visuell wahrnehmbar sind. Zwar berichten Versuchspersonen häufig von visuellen mentalen Bildern, wenn sie bspw. über die (vermutete) mechanische Funktionsweise von Systemen nachdenken. Diese Ableitung einer möglichen Funktionsweise erfordert allerdings mehr Informationen, als durch die bloße Inspektion eines mentalen Bildes zu gewinnen wären. Daher scheint es sich bei dieser visuellen Vorstellung um ein Epiphänomen bei der Anwendung von mentalen Modellen zu handeln, das unter gewissen Umständen sogar den Prozess des

Schlussfolgern behindern und verlangsamen kann (*visual impedance effect*; Knauff & Johnson-Laird, 2002; Knauff, 2006, S. 147). Hierzu gibt es allerdings divergierende Literaturmeinungen (vgl. Gottschling, 2006, S. 231).

Ferner sind die bisher betrachteten individuellen von interindividuellen mentalen Modellen abzugrenzen, die aus soziologischer Perspektive ein wesentliches Mittel zur intersubjektiven Verständigung darstellen. Von geteilten, gemeinsamen oder kollektiven mentalen Modellen wird gesprochen, wenn unterschiedliche Individuen über ähnliche mentale Modelle verfügen (Denzau & North, 1994). In Teams gewährleisten geteilte mentale Modelle, etwa von der Aufgabe des Teams oder dem Teamgefüge und der jeweiligen Austauschprozesse, die effektive Koordination der Aktivitäten innerhalb des Teams und begünstigen damit die Erreichung der gemeinsamen Ziele (Matthews, Panganiban, Lin, Long & Schwing, 2021, S. 62; Scheutz, DeLoach & Adams, 2017; Schuster et al., 2011; Talone, Phillips, Ososky & Jentsch, 2016). Kollektive mentale Modelle lassen sich als Instrumente der interpersonellen Verständigung begreifen, die in Summe eine gesellschaftliche Denkweise manifestieren können und die sich extern in der Struktur von Institutionen widerspiegeln. Somit erlaubt die Betrachtung des institutionellen Gefüges einer Gesellschaft Rückschlüsse auf die in ihr vorherrschenden mentalen Modelle. Während kollektive mentale Modelle die äußere Welt innerhalb der mentalen Repräsentation strukturieren, leisten das Institutionen analog in der externen Welt (Denzau & North, 1994). Damit bilden kollektive mentale Modelle zugleich die Grundlage für die soziale Legitimation von Institutionen. Wenn die mentalen Modelle der Institutionen von deren realen Ausgestaltung abweichen, kommt es zu einem Vertrauensverlust (Beckmann, Mackenbrock, Pies & Sardison, 2004).

Mentale Modelle werden im Rahmen dieser Arbeit als dynamische Repräsentationen verstanden, die auch abstrakte, nicht-visuelle Elemente enthalten, nicht an eine bestimmte Modalität gebunden sind und insbesondere den Ausgangspunkt für Simulationen zur Erklärung und Vorhersage der Aktionen technischer Systeme bilden (Gottschling, 2006, S. 213f.). Letztlich handelt es sich bei diesem Konzept der mentalen Modelle selbst wiederum um ein mentales Modell zur Erklärung menschlicher Denkvorgänge.

3.1.4 Empirische Zugänge

Da sich mentale Modelle als kognitive Konstrukte einer direkten Beobachtung oder Messung entziehen, erweist sich deren empirische Untersuchung als herausfordernd (Jones et al., 2011, S. 2; Rouse et al., 1992, S. 1304; Seel et al., 2009, S. 27). Während inzwischen auch neurowissenschaftliche Erkenntnisse das Konzept der mentalen Modelle stützen, gilt die Verhaltensforschung als paradigmatischer Ansatz zur Ergründung mentaler Modelle (Seel, 2017, S. 942). Eine vollständige Offenlegung der jeweiligen mentalen Modelle ist damit allerdings unmöglich (Rouse & Morris, 1986). Denn eine Untersuchung durch Dritte erfordert zunächst die prinzipielle Bereitschaft und die Fähigkeit der untersuchten

Personen, ihre Modelle zu externalisieren und zu kommunizieren (Seel et al., 2009, S. 27). Wenn sich Personen ihrer mentalen Modelle bewusst sind und sie diese verbalisieren können, wird von expliziten mentalen Modellen gesprochen, wohingegen implizite mentale Modelle zwar möglicherweise die Basis für Schlussfolgerungen bilden, aber sich als implizites Wissen (*tacit knowledge*) nicht explizit kommunizieren lassen (Seel, 2017; Stagers & Norcio, 1993). Eine besondere Herausforderung stellt die Untersuchung von Veränderungsprozessen mentaler Modelle dar, da dies wiederholte und vergleichbare Messungen des mentalen Modells zu verschiedenen Zeitpunkten erfordert (Seel et al., 2009, S. 27). Die Erhebung der lernabhängigen Weiterentwicklung mentaler Modelle gilt als bisher ungelöstes methodologisches Problem (Ifenthaler & Seel, 2012, S. 2034).

In den verschiedenen Forschungsdisziplinen kommen zur Untersuchung mentaler Modelle typischerweise unterschiedliche Erhebungsmethoden zum Einsatz (Jones et al., 2014). Interviewprozeduren lassen sich entlang zweier Dimensionen unterscheiden, nämlich erstens anhand situativer *versus* nicht-situativer und zweitens anhand verbal- *versus* visuell-basierter Methoden (Jones et al., 2014). Folgende Messmethoden werden in der Literatur besprochen:

1. *Methode des lauten Denkens* zur Generierung verbaler Protokolle bei der Systembenutzung (Preim & Dachselt, 2010, S. 96) sowie rezeptive Interviews und *teach-back*-Prozeduren (Seel, 2001, S. 415)
2. *Zeichen-Methode*, um einen visuellen Eindruck des mentalen Modells in Skizzenform zu erhalten (Bilotta, Gabriele, Servidio & Tavernise, 2007, S. 1; Ososky, Phillips, Schuster & Jentsch, 2013, S. 1300). Hierbei ist einschränkend zu berücksichtigen, dass sich mit diesem Verfahren primär visuelle Aspekte des mentalen Modells externalisieren lassen.
3. *Kausale Diagramme* als Möglichkeit, um Annahmen über kausale Zusammenhänge in einem beschriebenen System zu externalisieren (Seel, 2001, S. 415).
4. *Beobachtung der Systembenutzung* und Ableitung der Modelle aus Nutzer:innenbeschreibungen und -erwartungen bezüglich des Systemverhaltens (Bilotta et al., 2007, S. 1).
5. *Linguistische Analyse* der Äußerungen der Systembenutzenden. Stärker interaktionale Sprachmerkmale weisen bspw. auf eine Repräsentation eines technischen Systems als Interaktionspartner:in hin (Fischer, Lohan & Foth, 2012).

Die Instabilität und Kontextualität mentaler Modelle lässt vermuten, dass diese durch den Erhebungsprozess selbst beeinflusst werden. So gehen Jones et al. (2014, S. 5) davon aus, dass Personen eher generelle, grundsätzliche und langzeitbezogene Aspekte ihrer mentalen Modelle preisgeben, wenn die Befragung in ihrer gewohnten Umgebung stattfindet. Im Gegensatz dazu äußern Personen mutmaßlich mehr durch die Situation hervorgerufene und spezifische Elemente, wenn die Befragung in der konkreten Situation stattfindet, auf die sich das Modell bezieht, also z. B. in der Gegenwart eines Roboters (Jones et al., 2014,

S. 5). Eine beeinflussungsfreie Untersuchung mentaler Modelle ist daher unmöglich. Alternativ ist die mögliche Verzerrung der Datengrundlage durch die Befragungssituation vom Forschenden zu reflektieren. In diesem Zusammenhang ist zu unterscheiden, ob das Erkenntnisinteresse in einem Abbild des vorherrschenden mentalen Modells oder in dessen situationsbezogener Veränderung besteht. So schlagen Talone et al. (2016) vor, mentale Modelle vor der Konfrontation mit einem Roboter als *baseline* zu erheben und mit einer zweiten Messung nach einer Interaktion zu vergleichen. Bei einer solchen vergleichenden Erhebung lässt sich der Einfluss der Erhebungsmethode zumindest reduzieren.

3.1.5 Relevanz im MRI-Kontext

Im Rahmen der Mensch-Technik-Interaktion repräsentieren mentale Modelle die jeweiligen technischen Artefakte und bieten dem menschlichen Gegenüber damit bestenfalls eine adäquate Beschreibung des technischen Systems, eine Erklärung für dessen Funktionsweise sowie eine akkurate Vorhersagemöglichkeit über zukünftige Aktionen bzw. Reaktionen des Systems. Entsprechend werden mentale Modelle bereits seit den 1990er-Jahren (unter diesem oder ähnlichen Termini) im Bereich der Mensch-Computer-Interaktion und insbesondere der Nutzungserfahrung (*user experience*) diskutiert (Seel, 2017, S. 940; Stagers & Norcio, 1993, S. 587). Aus Sicht der Entwickler:innen technischer Systeme besteht das Ziel darin, eine möglichst effektive und fehlerfreie Benutzung derselben zu gewährleisten, bei der die Nutzer:innen über ein mentales Modell verfügen, das mit dem tatsächlichen Systemverhalten korrespondiert (Preim & Dachsel, 2010, S. 94f.).

Auch im MRI-Kontext fordern inzwischen einige Autor:innen die explizite Berücksichtigung mentaler Modelle (Devitt, 2018, S. 174; Ososky, 2013). Die Relevanz mentaler Modelle ergibt sich in diesem Kontext aus der Annahme, dass diese die konkreten Erwartungen einer Person in Hinblick auf die Beschaffenheit, die Fähigkeiten und die Aktionen eines Roboters innerhalb eines Interaktionsszenarios prägen und damit die Grundlage für das Verständnis über die Arbeitsweise des beteiligten technischen Systems bilden (Kiesler, 2005, S. 730). Obwohl die Forschung zu mentalen Modellen durch Kognitionswissenschaftler:innen über eine lange Tradition verfügt (Jones et al., 2011), wurde das Konzept der mentalen Modelle sowie deren Beeinflussbarkeit und Auswirkungen in MRI-Szenarien nur in vereinzelt Studien und häufig nur in oberflächlicher Form einbezogen (Ososky, 2013, S. 3). In einer jüngeren MRI-Studie definieren de Visser et al. (2020, S. 460) mentale Modelle bspw. sehr allgemein als „interne Repräsentation in der Vorstellung eines Akteurs über die Eigenschaften eines anderen Akteurs“ (*an internal representation in the mind of one actor about the characteristics of another actor*). Einige Studien beziehen sich in ihren Erläuterungen auf verwandte kognitive Konzepte wie Schemata (L. J. Baker, Hymel & Levin, 2018), Konzeptualisierungen (Fischer, 2011), kognitive Frames (Sauppé & Mutlu, 2015) oder das Priming von Erwartungen durch abrufbare soziale Skripte (A. Edwards et

al., 2019). Hierbei ist häufig unklar, inwiefern diese Konzepte im jeweiligen Studienkontext auch als mentale Modelle verstanden werden können bzw. welche Gemeinsamkeiten und Abgrenzungen bestehen.

Die Entwicklung mentaler Modelle von einem Roboter kann bereits vor der tatsächlichen Interaktion mit demselben beginnen, weil bereits zu diesem Zeitpunkt ein Prozess der Sinngebung (*sense-making*) einsetzt (Mutlu & Forlizzi, 2008). Diese sog. mentalen Modelle *a priori* prägen anschließend die Wahrnehmung, die Bewertung und das Verhalten während der Interaktion (Charalambous et al., 2016a; Kiesler, 2005; Komatsu, Kurosawa & Yamada, 2012; Ososky et al., 2013; Phillips et al., 2011; Syrdal, Dautenhahn, Koay, Walters & Otero, 2010; Weis & Wiese, 2020). Solange keine tatsächlichen Erfahrungswerte über Roboter zur Verfügung stehen, liegt die Annahme nahe, dass diese mentalen Modelle aus Übertragungen der als Schablone verfügbaren Modelle von als ähnlich empfundenen Entitäten entstehen. Sowohl andere Maschinen, seriöse Medienberichte als auch *science fiction*-Literatur sowie Filme aus der Popkultur bieten dabei ein variantenreiches Spektrum an Vorlagen (Ososky et al., 2013). Entsprechend wiesen die Zeichnungen eines fiktiven Haushaltsroboters der Teilnehmenden einer Studie eine hohe Ähnlichkeit mit filmischen Roboter-Darstellungen auf. Die entstandenen Skizzen ließen sich entweder als menschen- oder maschinenähnlich klassifizieren, was auf die Anwendung zweier vorhandener, aber elementar unterschiedlicher mentaler Modelle hindeutet (Broadbent, Lee, Stafford, Kuo & MacDonald, 2011). In einer weiteren Studie mit einem Transportroboter in einem Krankenhaus zeigten sich in Interviews vier verschiedene Repräsentationen des Roboters: als eine Art außerirdisches Wesen, als Maschine, als Mitarbeitende:r und als Kolleg:in. Dabei fielen diese Repräsentationen mit zunehmender Gewöhnung und Nutzungserfahrung menschenähnlicher und kollegialer aus (Ljungblad, Kotrbova, Jacobsson, Cramer & Niechwiadowicz, 2012; Sauppé & Mutlu, 2015). Ferner identifizierte Mabry (2016) drei gängige Narrative, die die Hierarchie zwischen Menschen und sozialen Robotern umschreiben, nämlich mittels eines Roboters als Untergebene:r des Besitzenden, der aufbegehrt und Widerstand leistet, eines Roboters als gefügte:r Diener:in, und eines Roboters als kinderähnliches Wesen, das nach der Liebe und der Gemeinschaft des Besitzenden begehrt (Mabry, 2016).

Besonders brisant ist der Entstehungsprozess eines neuen mentalen Modells, wenn verschiedene Hinweisreize des *modellandum* wahrgenommen werden, die sich mit verschiedenen bestehenden mentalen Modellen in Verbindung bringen lassen. So stellt sich bspw. bei hybriden Cobots die Frage, inwiefern ein kohärentes mentales Modell von einem bekanntermaßen unbelebten Roboter erzeugt werden kann, der sich allerdings menschenähnlich und/oder sozial verhält (Kiesler & Goetz, 2002). Offensichtlich ist hier sowohl ein Rückgriff auf ein mentales Modell von Menschen als auch auf ein solches von Maschinen möglich. Złotowski et al. (2014) gehen in diesem Zusammenhang davon aus, dass Menschen in Ermangelung passender mentaler Modelle intuitiv dazu tendieren,

anthropomorphe mentale Modelle von real unbelebten Objekten zu kreieren. Anthropomorphe mentale Modelle stehen demnach als universelle Rückfallmöglichkeit zur Verfügung, wenn passende mentale Modelle für gewisse Entitäten fehlen (Złotowski et al., 2014). Das verdeutlicht zugleich die Schnittmengen zwischen dem Konzept der mentalen Modelle und dem Phänomen des Anthropomorphismus (vgl. Kapitel 3.2).

Wie bereits in Kapitel 3.1.2 beschrieben, können sich auch unvollständige mentale Modelle als zweckdienliche und effektive Instrumente zur Interaktion mit komplexen Systemen erweisen (Ososky, 2013, S. 11; Preim & Dachsel, 2010, S. 96). Insbesondere mentale Modelle *a priori* bergen allerdings aufgrund der eingeschränkten Informationsbasis die Gefahr, zu faktisch unzutreffenden Erwartungen und Rückschlüssen über die Fähigkeiten und Einschränkungen eines Roboters zu führen, die die prinzipielle Ungenauigkeit von Modellen übersteigen (Phillips, Schaefer, Billings, Jentsch & Hancock, 2016; Talone et al., 2016). Es ist davon auszugehen, dass solche mentalen Modelle übermäßig stark durch sichtbare oberflächliche Robotereigenschaften wie dessen Design oder Bewegungsform sowie durch selektiv vermittelte Informationen oder durch Darstellungen in der Popkultur geprägt sind (Ososky et al., 2013; Talone et al., 2016). Stehen die aus dem mentalen Modell abgeleiteten Erwartungen zu realen Erfahrungen in Widerspruch, weil z. B. einem Roboter spezifisch menschliche Fähigkeiten zugeschrieben wurden, kann sich dies negativ auf die Zusammenarbeit in hybriden Mensch-Roboter-Teams auswirken (Laue, 2017; Phillips et al., 2011; Talone et al., 2016). So zeigten die Versuchsteilnehmenden in einer Studie von Broadbent et al. (2011) negative physiologische und psychologische Reaktionen auf einen Roboter, der ihrem mentalen Modell widersprach. Ferner scheint die Übertragung mentaler Modelle von anderen Entitäten auf einen Roboter einen Vergleichsrahmen für die Beurteilung dieses Roboters festzulegen, wie Syrdal et al. (2010) im Zuge einer qualitativen, explorativen Studie mit einem zoomorphen Roboter vermuten. Die hundeähnliche Repräsentation des Roboters erlaubte zwar exakte Vorhersagen von dessen Aktionen, ließ den Roboter allerdings im Vergleich zu einem realen Hund weniger nützlich wirken. Eine Versuchsperson hob explizit hervor, dass es unmöglich sei, zu dem Roboter eine Beziehung wie zu einem Hund aufzubauen (Syrdal et al., 2010, S. 644).

In Bezug auf die Einführung eines Cobots am industriellen Arbeitsplatz spielen mentale Modelle hinsichtlich des Vertrauens von Mitarbeitenden und der Effizienz der Interaktion eine zentrale Rolle (Wilkison, Fisk & Rogers, 2007). In diesem Zusammenhang besteht weiterer Forschungsbedarf, um die Determinanten und Auswirkungen bestehender mentaler Modelle auf den Umgang mit Robotern detaillierter zu ergründen (S. Lee et al., 2005; Ososky et al., 2013; Powers & Kiesler, 2006), insbesondere angesichts neuer Generationen hybrider und interaktiver Roboter (Sheridan, 2016). Im Rahmen der nachfolgenden Theoriekapitel werden die Auswirkungen von mentalen Modellen auf Vermenschlichungsvorgänge und Vertrauensbildungsprozesse sowie die Beeinflussbarkeit dieser Prozesse durch sprachliches Framing genauer exploriert und spezifiziert.

3.2 Anthropomorphismus

„There is an universal tendency amongst mankind to conceive all beings like themselves, and to transfer to every object those qualities with which they are familiarly acquainted, and of which they are intimately conscious. We find human faces in the moon, armies in the clouds; and by a natural propensity, if not corrected by experience and reflection, ascribe malice and good will to everything that hurts or pleases us.”

– Hume, 1757/1994, S. 29 –

3.2.1 Begriffsbestimmung

Das Zitat von David Hume dokumentiert exemplarisch, wie lange die Überlegungen zum Anthropomorphismus zurückreichen und welche prominenten Denker:innen sich mit diesem Phänomen auseinandergesetzt haben (Epley et al., 2007). Der Begriff des Anthropomorphismus bezieht sich auf die menschliche Neigung zur Vermenschlichung (Anthropomorphisierung), d. h. zur Zuschreibung menschlicher oder menschenähnlicher Eigenschaften an unbelebte Entitäten wie Roboter (B. R. Duffy, 2003; Epley et al., 2007). In diesem Sinne basiert Vermenschlichung auf einem generalisierenden induktiven Schluss, bei dem das mentale Modell von Menschen bzw. vom eigenen Selbst (Einzelfall) auf eine Klasse an Entitäten (Verallgemeinerung) projiziert wird (Airenti, 2015; Epley, Waytz, Akalis & Cacioppo, 2008; Wiese, Metta & Wykowska, 2017). Diese Definition kann weitestgehend als Konsens angesehen werden, wenngleich einige Studien den Vermenschlichungsbegriff nur oberflächlich einführen und damit zu Unschärfen in der Verwendung beitragen (Marquardt, 2017, S. 24). Solche Unschärfen werden bspw. dadurch begünstigt, dass manche Autor:innen den Begriff der Anthropomorphisierung mitunter zur Beschreibung der anthropomorphen Gestaltung eines Roboters verwenden. So definieren bspw. Kuz, Mayer, Müller und Schlick (2017) Anthropomorphismus als die Simulation menschlicher Eigenschaften durch nicht-menschliche Entitäten. Allerdings stellt das anthropomorphe Design einer Entität weder ein hinreichendes noch ein notwendiges Kriterium dafür dar, dass menschliche Beobachtende diese Entität in ihren mentalen Modellen menschenähnlich repräsentieren. Schließlich führt einerseits eine menschenähnliche Gestaltung nicht zwangsweise zur Anthropomorphisierung, z. B. dann nicht, wenn dezidiert maschinelle mentale Modelle von der Entität vorliegen, und andererseits vermenschlichen Personen häufig auch solche Objekte, die nicht dezidiert anthropomorph gestaltet wurden (Marquardt, 2017, S. 9). Das Phänomen des Anthropomorphismus lässt sich daher keineswegs auf rein objektive, beobachtungsunabhängige Eigenschaften einer Entität reduzieren, sondern findet in der Interaktion zwischen äußerer Erscheinung und mentalem Deutungsrahmen statt (Persson, Laakso & Lonnqvist, 2000). Zur Abgrenzung wird in der vorliegenden Dissertation von einem anthropomorphen Design gesprochen, um den Vorgang der menschenähnlichen Ausgestaltung eines Roboters zu umschreiben. Dahingegen bezeichnet

Anthropomorphisierung den psychischen Vorgang des mentalen Zuschreibens menschlicher Attribute an eine unbelebte Entität, wobei die Wahrnehmung eines anthropomorphen Designs diese Zuschreibung begünstigen oder auslösen kann (Fink, 2012; Lemaignan, Fink, Dillenbourg & Braboszcz, 2014). Tabelle 5 stellt ergänzend verschiedene typische Merkmale des anthropomorphen Designs dar.

Tabelle 5: Typische Merkmale des anthropomorphen Designs.

Kategorie (Fink, 2012, S. 200)	Merkmalstyp (Fogg, 2004, S. 92)	Beispiele
Äußere Gestalt	Physikalische Reize	Augen, Körper oder Gesicht
(Sozial-)Verhalten	Psychologische Reize	Humor, Präferenzen oder Gefühle
Interaktion/ Kommunikation	Sprachliche Reize	Verstehen und Äußern menschlicher Sprache sowie die Verwendung interaktionaler Sprachmerkmale
(Sozial-)Verhalten	Soziale Dynamiken	Kooperation, Lob oder Reziprozität
(Sozial-)Verhalten	Soziale Rolle	Konkurrent:in, Weisungsbefugte:r

3.2.2 Charakterisierung und Differenzierung

3.2.2.1 Zweistufiges Modell des Menschseins

Nach dieser ersten Bestimmung und Abgrenzung des Begriffs des Anthropomorphismus soll dieser nachfolgend einer tiefergehenden Charakterisierung und einer Ausdifferenzierung in seine verschiedenen Spielarten unterzogen werden. Der Anthropomorphismus war seit jeher Gegenstand sowohl psychologischer Forschung als auch philosophischer Debatten, da diese menschliche Neigung die Fragen nach der *conditio humana* berührt (Zawieska, Duffy & Spronska, 2012). Entsprechend finden sich in psychologischen Theorien auch Rückgriffe auf und Überschneidungen mit philosophischen Denkansätzen, wie bspw. mit dem sog. *Ansatz der intentionalen Haltung (intentional stance; Dennett, 1971)*. Gemäß diesem Ansatz erfüllt die Zuschreibung von Handlungsabsichten die Funktion, Vorhersagen über die künftigen Aktionen eines Gegenübers abzuleiten (Remmers, 2020, S. 66). Übereinstimmend bestimmt auch die anwendungsorientierte psychologische Forschung den Nutzen des Anthropomorphismus darin, die Aktionen einer unbekannt

Entitäten zu rationalisieren und zu interpretieren, um damit innerhalb der gewohnten sozialen Deutungsmuster mit derselben interagieren zu können (Bergamasco & Herr, 2016, S. 1900f.; Compagna et al., 2016; Darling, 2016; B. R. Duffy, 2003; Hinds et al., 2004; Phillips et al., 2011). Diese Funktionsbestimmung verdeutlicht, dass eine pauschale Abwertung von vermenschlichenden Zuschreibungen als kognitive Fehlleistung zu kurz greift. Vielmehr stellt der Anthropomorphismus einen effizienten Mechanismus zur Interaktion mit der Umwelt dar (Damiano & Dumouchel, 2018; Loh, 2019, S. 113) und verweist dabei erneut auf das Spannungsfeld zwischen deskriptiver und präskriptiver Funktion von mentalen Modellen. Wenngleich vermenschlichende Zuschreibungen mit einem unzutreffenden mentalen *Modell* von einer unbelebten Entität einhergehen, können sie ein geeignetes mentales *Modell* für die Interaktion mit derselben darstellen.

Anthropomorphismus umfasst allerdings mehr als die bloße Zuschreibung von Handlungsabsichten aus Gründen der Vorhersagbarkeit (Schellen & Wykowska, 2019, S. 2) und reicht insbesondere über die reine Beschreibung von Verhaltensweisen hinaus (Epley et al., 2008). Das beobachtbare Verhalten eines Tieres als aggressiv zu bezeichnen, würde bspw. keine Vermenschlichung darstellen. Ferner geht Anthropomorphismus über die reine Zuschreibung von Lebendigkeit hinaus, da Lebendigkeit keine dem Menschen exklusiv vorbehaltene Eigenschaft darstellt (Epley et al., 2008). Die Neigung zur Zuschreibung von Lebendigkeit und/oder Bewusstsein wird als Animismus bezeichnet und lässt sich sowohl bei einigen traditionellen Völkern sowie entwicklungspsychologisch bei Kindern in ausgeprägter Form vorfinden (Samuel, 2019). Eine Vermenschlichung i. e. S. läge demgegenüber erst vor, wenn typisch menschliche mentale Zustände und Eigenschaften zugeschrieben werden, die sich nicht direkt äußerlich beobachten lassen (Waytz et al., 2014). Animismus gilt demnach als notwendige, aber nicht als hinreichende Bedingung für Anthropomorphismus (Zawieska et al., 2012)

An diese Abgrenzung schließt die Frage an, welche Eigenschaften einer Entität zugeschrieben werden müssten, damit von einer Zuschreibung von Menschlichkeit gesprochen werden kann. Gray, Gray und Wegner (2007) verweisen darauf, dass grundsätzlich zwei Merkmale als spezifisch menschlich gelten, nämlich die Handlungsfähigkeit als (autonome:r und rationale:r) Akteur:in (*agency*; auch: Akteur:inschaft) und die Empfindungsfähigkeit (*experience*). Ein ähnliches zweistufiges Konzept findet sich in der Literatur zur Dehumanisierung, also dem Prozess des Absprechens der Menschlichkeit (Haslam, 2006). Demnach gibt es einerseits Charakteristika, die einzigartig menschlich sind (*uniquely human*, UH), und andererseits solche, die die Zugehörigkeit zur weiter gefassten menschlichen Natur (*human nature*, HN) bestimmen. Die Empfindungsfähigkeit gilt dabei als Merkmal der menschlichen Natur, das auch bei Tieren in mehr oder weniger stark ausgeprägter Form vorhanden ist, wohingegen die Handlungsfähigkeit bzw. Akteur:inschaft als einzigartig menschlich, also nur Menschen vorbehalten, gilt. Złotowski et al. (2014) und Złotowski, Sumioka, Bartneck, Nishio und Ishiguro (2017) konzeptualisieren nun

Anthropomorphismus als eine Umkehrung des Prozesses der Dehumanisierung und beziehen damit das zweistufige Modell des Menschseins auf den Anthropomorphismus. Gemäß diesem Ansatz können nur Entitäten, die der menschlichen Wesensart angehören, also über Empfindungsfähigkeit verfügen, auch über einzigartig menschliche Merkmale wie die Handlungsfähigkeit bzw. Akteur:inschaft verfügen. Das Vorhandensein von HN-Merkmalen wäre damit eine notwendige Bedingung dafür, dass zusätzlich UH-Merkmale vorliegen können. Spräche man einem Menschen seine Zugehörigkeit zur menschlichen Wesensart ab (HN-Merkmale), erhielte dieser den Status einer Maschine, wohingegen das bloße Inabredestellen der Handlungsfähigkeit (UH-Merkmale) einen Menschen auf die Stufe eines nicht-menschlichen Tieres stellte. Tabelle 6 verdeutlicht diese Zusammenhänge auf Basis eines zweistufigen Modells des Menschseins.

Tabelle 6: Zweistufiges Modell des Menschseins und der Vermenschlichung auf Basis von Empfindungs- und Handlungsfähigkeit nach Złotowski et al. (2017).

	Empfindungsfähigkeit	Handlungsfähigkeit
Menschliches Lebewesen	vorhanden	vorhanden
Nicht-menschliches Lebewesen	vorhanden	nicht vorhanden
Unbelebte technische Artefakte	nicht vorhanden	nicht vorhanden

Einschränkend ist zu erwähnen, dass bis dato keine empirische Evidenz für diesen zweistufigen Prozess der Vermenschlichung vorliegt und die Einordnung von bestimmten Merkmalen als spezifisch menschlich zeitabhängig und im Kontrast zur verfügbaren Technologie jeweils unterschiedlich vorgenommen wird (Heßler, 2015). Besonders bemerkenswert ist in diesem Kontext, dass empirische Studien entgegen der vorgenommenen Einordnung zeigten, dass die wahrgenommene Empfindungsfähigkeit eines Roboters einen stärkeren Einfluss darauf hatte, ob Versuchspersonen diesen Roboter als menschlich wahrnahmen (Waytz & Norton, 2014). Möglicherweise erscheint in einer Zeit des zunehmenden Aufkommens von Maschinen mit ausgeprägten kognitiven Fähigkeiten die Empfindungsfähigkeit als herausragendes menschliches Merkmal. Diese These wird durch den Befund bekräftigt, dass Mitarbeitende eher dazu bereit sind, die Ersetzung ihrer kognitiven Tätigkeiten durch Automatisierungstechnologie zu akzeptieren anstatt die Substitution ihrer emotionalen Aufgaben, die sie womöglich stärker mit ihrem Selbstverständnis als Menschen verbinden (Waytz & Norton, 2014).

In der anwendungsorientierten Forschung wird häufig von Anthropomorphisierung gesprochen, ohne genauer zu spezifizieren, welche Zuschreibungen inhaltlich damit verbunden sind. Einen stärker ausdifferenzierten Ansatz zur qualitativen Unterscheidung des Inhalts verschiedener anthropomorpher Zuschreibungen verfolgen Persson et al. (2000) in einem

hierarchischen fünfstufigen Modell. Dieses bringt typische Arten von Zuschreibungen in eine hierarchische Ordnung aufsteigender Komplexität. Auf der niedrigsten Komplexitätsebene I befinden sich spontane und basale Kategorisierungen u. a. in *lebendig* und *nicht lebendig*, die v. a. auf Basis visueller Reize als Bestandteil des anthropomorphen Designs erfolgen. Auf den höheren Komplexitätsebenen II und III folgen Zuschreibungen auf Basis einer primitiven Psychologie und auf Basis der Alltagspsychologie. Auf den beiden höchsten Stufen IV und V werden komplexe Charaktereigenschaften und soziale Rollen zugeschrieben, die ganze Bündel von Annahmen implizieren (Persson et al., 2000). Vor dem Hintergrund der vorgenannten Unterscheidungen von Merkmalen des Menschseins erscheint ein solcher Ansatz gewinnbringend. Einschränkend ist allerdings zu erwähnen, dass dieser Ansatz noch unzureichend ausgearbeitet und empirisch untermauert ist und bisher kaum rezipiert wurde (Marquardt, 2017, S. 12).

3.2.2.2 Starker und schwacher Anthropomorphismus

Anthropomorphe Zuschreibungen entstehen nicht notwendigerweise auf Basis vernünftiger oder logisch begründbarer Überlegungen. Vielmehr haben vermenschlichende Zuschreibungen den Stellenwert eines Glaubenssatzes oder einer Meinung, die zu einem unterschiedlich starken Grad vertreten wird (Epley et al., 2008). Entsprechend unterscheiden Epley et al. (2007, S. 867) in diesem Zusammenhang explizit zwischen starkem und schwachem Anthropomorphismus, wobei sich diese beiden Spielarten nicht prinzipiell, sondern nur graduell unterscheiden. Bei der schwachen Form sprechen Menschen zwar metaphorisch von einer nichtmenschlichen Entität so, *als ob* diese menschliche Eigenschaften besitze.¹ In Abgrenzung zur starken Form vertreten die entsprechenden Personen allerdings nicht explizit die Auffassung, dass dies tatsächlich so ist. Folglich würden sie sich dagegen verwehren, dieser Entität einen menschenähnlichen Status zuzuschreiben. In der Praxis lassen konkrete vermenschlichende Aussagen über einen Roboter selten Rückschlüsse darauf zu, ob es sich um eine schwache oder starke Form der Vermenschlichung handelt (Zawieska et al., 2012), solange keine expliziten Rückfragen nach dem Status der Entität gestellt werden. Starker Anthropomorphismus findet sich typischerweise bei der religiösen Rede über Gott. Besonders in diesem Kontext lässt sich ein objektiver Wahrheitsgehalt solcher Zuschreibungen zumindest im Rahmen einer psychologischen Theorie nicht bestimmen, sondern fällt vielmehr in den Kernbereich der Philosophie und ist dort vor dem Hintergrund verschiedener Strömungen unterschiedlich zu beurteilen (Epley et al., 2008). In Bezug auf die Weiterentwicklung von Robotern lässt sich in diesem Zusammenhang ferner zwischen einer schwachen und einer starken ontologischen Konzeption unterscheiden. Vertreter:innen der schwachen ontologischen Konzeption gehen davon aus, dass anthropomorphe Roboter zunehmend menschlich zu werden *scheinen* (*appear to become*

¹ Für weiterführende philosophische Analysen zu verschiedenen Formen und Bedeutungen dieser *as if*-Aussagen siehe Gutmann und Knifka (2015, S. 70ff.).

human). Anhänger der starken ontologischen Konzeption gehen hingegen davon aus, dass humanoide Roboter tatsächlich menschlich *werden* (*actually become human*), also denselben ontologischen Status wie Menschen erlangen. Je nach vertretener ontologischer Konzeption lassen sich vermenschlichende Zuschreibungen demnach entweder als bloße Metaphern oder als ontologisch zutreffende Beschreibungen auffassen (Zawieska et al., 2012, S. 80).

Bei der vorgenommenen Unterscheidung in starke und schwache Formen des Anthropomorphismus ist in Hinblick auf dessen Konsequenzen die Feststellung wichtig, dass sowohl die schwache als auch die starke Form der Vermenschlichung erhebliche Auswirkungen auf das Verhalten der jeweiligen Person haben kann, da sich dieses an den sprachlichen Metaphern orientiert (Epley et al., 2007). Wenngleich vermenschlichende Aussagen nicht zwangsläufig als deskriptive Feststellungen i. e. S., sondern z. B. auch als pragmatische Umschreibungen verstanden werden können, trägt die vermenschlichende Rede über Technologie dazu bei, dass sich mit der Zeit konsistente vermenschlichende Annahmen ausbilden (Damiano & Dumouchel, 2018). Ferner kommunizieren vermenschlichende Zuschreibungen über die verwendeten Termini stets Erwartungen an die Technologie, die im öffentlichen Diskurs zu kollektiven Erwartungshaltungen avancieren können (B. R. Duffy, 2003; Marquardt, 2017, S. 9).

3.2.3 Theoretische Ansätze

In diesem Kapitel wird zunächst die *media equation*-These (Kapitel 3.2.3.1) beschrieben, die das Phänomen der (schwachen) Vermenschlichung stark in den Fokus der Forschung und der öffentlichen Diskussion rückte. Obwohl die Autoren ihre Ergebnisse ursprünglich nicht aus dem Blickwinkel der Vermenschlichung, sondern des instinktiven Sozialverhaltens interpretierten, lassen sich wesentliche Schnittmengen identifizieren. Diese zeigen sich besonders vor dem Hintergrund der *Theorie des dualen Anthropomorphismus* (Kapitel 3.2.3.2). Anschließend wird mit der *Drei-Faktoren-Theorie* (Kapitel 3.2.3.3) ein psychologischer Ansatz präsentiert, der das Phänomen des Anthropomorphismus derzeit am umfassendsten beleuchtet. Das *Modell der Frame-Selektion* (Kapitel 3.2.3.4) ergänzt diese Betrachtungen um eine soziologische Sichtweise, die bereits eine deutlichere Nähe zum Konzept des Framings aufweist.

3.2.3.1 Media Equation-These

In ihren viel rezipierten Forschungsarbeiten untersuchten Reeves und Nass (1998) das Verhalten von Menschen gegenüber verschiedenen Medien wie Computern, Fernsehern und modernen Kommunikationstechnologien, um zu ergründen, inwiefern sich die Interaktionsstrukturen zwischen Menschen in gleicher oder ähnlicher Weise beim Kontakt mit Medien zeigen (Mou & Xu, 2017). Die Herangehensweise von Reeves & Nass (1998) bestand

darin, bekannte sozialpsychologische Experimente zu replizieren, dabei allerdings das menschliche Gegenüber aus den Originalstudien durch ein technisches Medium zu ersetzen. Dabei ergaben sich in erstaunlichem Ausmaß übereinstimmende Ergebnisse zwischen den Original- und den Replikationsstudien. So zeigte sich bspw., dass Menschen stereotypische soziale Kategorien wie das Geschlecht, den ethnischen Hintergrund oder die Gruppenzugehörigkeit anwendeten, um je nach wahrgenommener Kategorie unterschiedlich auf Computer zu reagieren. Weibliche Computerstimmen wurden mit höherer Expertise in Themenbereichen wie Liebe und Partnerschaft assoziiert. Wenn das dargestellte Gesicht eines virtuellen Agenten optische Hinweise darauf lieferte, dass dieser denselben ethnischen Hintergrund wie die jeweilige Versuchsperson hat, wurde der virtuelle Agent u. a. als attraktiver, vertrauenswürdiger und intelligenter eingestuft (Nass & Moon, 2000). Darüber hinaus ergaben die Experimente, dass die Vergabe eines gemeinsamen Teamnamens oder das Tragen eines gleichfarbigen Teambändchens genügte, damit ein Computer als Teammitglied empfunden wurde. Die beteiligten Versuchspersonen orientierten daraufhin ihre Antworten stärker an denen des Computers und erachteten dessen Informationen als relevanter, empfanden die Präsentation der Informationen als freundlicher und zeigten stärkere Kooperationsgefühle (Reeves & Nass, 1998, S. 156ff.). Ferner offenbarten Versuchspersonen in verschiedenen Experimenten Merkmale eines komplexen Sozialverhaltens, wie bspw. ein höfliches Auftreten und die Anwendung des Prinzips der Reziprozität. So bewerteten die Versuchspersonen einen Computer positiver, wenn die entsprechenden Bewertungsfragen an diesem Computer selbst ausgefüllt werden mussten anstatt auf einem anderen Gerät. Das erinnert an den *experimenter bias*, der aus Studien mit einer menschlichen Versuchsleitung bekannt ist. Außerdem waren Versuchsteilnehmende eher dazu bereit, einem Computer bei einer Aufgabe zu helfen, wenn dieser sie zuvor bei einer anderen Aufgabe unterstützt hatte (vgl. als Überblick Broadbent, 2017, S. 640).

Auf Basis dieser Ergebnisse formulierten Reeves und Nass (1998) die These, dass der Umgang von Menschen mit Medien wie bspw. Computern dem mit realen, lebendigen Entitäten gleiche. Diese Gleichsetzung des menschlichen Verhaltens in Bezug auf medial vermittelte und tatsächlich anwesende Wesen bezeichneten Reeves und Nass (1998) als *media equation*. Dabei fassten sie das Spektrum möglicher Medien sehr breit und gingen über zur damaligen Zeit moderne Medien wie IMAX-Kinos oder Systeme zur Erzeugung einer virtuellen Realität (VR) hinaus. Die Autoren hoben dezidiert hervor, dass sich menschliches Sozialverhalten gleichermaßen in Bezug auf „Text auf Computern, ein computergesteuertes Heimkino, kleine und große TV-Geräte, Stimmen in multimedialen Lernvideos oder Bewegung innerhalb politischer Werbeanzeigen“ (Reeves & Nass, 1998, S. 252; Übersetzung des Verfassers) beobachten lässt. Spezifischer auf Computer bezogen leiteten sie daraus das *computers are social actors*-Paradigma (CASA-Paradigma) ab (Nass, Fogg & Moon, 1996; Reeves & Nass, 1998). Dieses Paradigma wurde in der Folge häufig rezipiert, um die Neigung von Menschen zu beschreiben, sich gegenüber Computern genauso wie gegenüber sozialen Wesen zu verhalten. Solche Reaktionen auf Medien sind den

Autoren zufolge in fundamentaler Weise sozial und natürlich und damit ein häufig anzutreffendes universelles und unbewusstes Phänomen, das leicht und ohne aufwändig produzierte Medien evozierbar und auch durch bewusste Reflexion kaum vermeidbar ist (Reeves & Nass, 1998). Menschen reagieren quasi instinktiv sozial, sobald ein Objekt auch nur den Anschein erweckt, lebendig zu sein (Brandstetter, 2017, S. 19). Vor dem Hintergrund einer interindividuell unterschiedlichen Disposition zur Vermenschlichung lässt sich allerdings nicht allgemeingültig festlegen, welche Merkmale ein Objekt mindestens aufweisen muss, um Sozialverhalten hervorzurufen (Waytz, Cacioppo & Epley, 2010). Gemäß der ursprünglichen Theorie ist die kognitive Einsicht, dass gewisse Entitäten wie bspw. Computer keine Persönlichkeit haben können, irrelevant für das Auftreten sozialer Verhaltensweisen. Die situative Wahrnehmung ist demnach wesentlich einflussreicher auf die menschliche Reaktion als die allgemeinen Überzeugungen bezüglich des Wesensstatus des jeweiligen Objekts (Reeves & Nass, 1998, S. 252ff.).

In der Literatur ist zwischenzeitlich sehr gut belegt, dass sogar technisch extrem simple, abstrakte und technomorphe Objekte ein elaboriertes Sozialverhalten hervorrufen können (Darling, 2017; Erel et al., 2018; Laue, 2017). Als historisches Beispiel dienen u. a. die *Braitenberg-Vehikel*, die technologisch lediglich auf einer einfachen Verschaltung zwischen zwei Lichtsensoren und zwei Antriebsmotoren basieren. Auf Beobachtende wirken deren Bewegungen so, als suchten bzw. mieden diese Fahrzeuge Lichtquellen, sodass sie die Vehikel mit psychologischem Vokabular beschreiben und ihnen ein komplexes, intentionales Verhalten zuschreiben (Braitenberg, 1987). Die Vermenschlichung manifestiert sich ferner in alltäglichen Aussagen wie bspw. „Er will heute wieder nicht!“ in Bezug auf einen nicht funktionierenden Computer, der nach *common sense*-Auffassung nicht über einen eigenen Willen verfügt (Phillips et al., 2011), oder in Aussagen, in denen das Gerät in zweiter Person direkt angesprochen wird (Coeckelbergh, 2010b). Mitunter haben Termini aus der Psychologie in Form anthropomorpher Metaphern Verwendung in den Computerwissenschaften gefunden, um Maschinen zu beschreiben. Umgekehrt halten fortwährend Begriffe aus der Computerwissenschaft als technomorphe Metaphern Einzug in die Alltagssprache zur Charakterisierung menschlicher Denkprozesse (Turkle, 2005, S. 2).

Als Ursache für die *media equation* vermuteten Reeves und Nass (1998) den Umstand, dass das menschliche Gehirn evolutionär daran gewöhnt sei, dass einerseits nur Menschen umfangreiches Sozialverhalten an den Tag legen und andererseits wahrgenommene Gegenstände reale Objekte darstellen, da bis vor wenigen Jahrzehnten keine virtuellen oder rein medial vermittelten Objekte existierten (Reeves & Nass, 1998, S. 12). Insofern verweisen die experimentellen Befunde nicht auf die Mächtigkeit der Medien im Sinne einer Technologiewürdigung, sondern auf die Einschränkungen des heuristischen menschlichen Denkens (Messaris, 2000). In dieser defizitären Sichtweise lässt sich die gedankenlose Anwendung eines sozialen Verhaltensmusters auch als fehlerhafte Übergeneralisierung verstehen (Fischer, 2011). Reeves und Nass (1998) verweisen in diesem Zusammenhang

mitunter darauf, wie simpel und undifferenziert die menschlichen Reaktionsmechanismen auf als solches beobachtetes Sozialverhalten seien. Kritiker betrachten diese vermeintlich einfachen Deutungen allerdings als Konsequenz der unterkomplexen Interpretationen der Autoren, die z. B. sehr vage bei der genaueren Bestimmung derjenigen Arten von Medien bleiben, auf die die *media equation* anwendbar ist (Dourish, 1996).

Ursprünglich wurde die *media equation*-These nicht mit Anthropomorphismus in Verbindung gebracht. Denn Nass und Moon (2000) definierten Anthropomorphisierung als *bewusste* Zuschreibung von Menschlichkeit an unbelebte Objekte, d. h. sie erachteten es als notwendig, dass Personen wirklich davon ausgehen, dass es sich bei dem Objekt um ein menschenähnliches Wesen handelt. Da alle Versuchsteilnehmenden in den Studien betonten, dass sie niemals sozial auf einen Computer reagieren würden, und gleichsam ihre zuvor gezeigten sozialen Verhaltensweisen verleugneten, schlossen Reeves und Nass (1998) eine so definierte Vermenschlichung als Erklärung für ihre experimentellen Beobachtungen im Bereich der *media equation*-These aus. Damit wendeten sich Nass & Moon (2000) allerdings explizit nur gegen die starke Form des Anthropomorphismus, während sich die CASA-These durchaus mit schwachem Anthropomorphismus erklären lässt, der für unterbewusste Zuschreibungen verantwortlich sein könnte (siehe auch Y. Kim & Sundar, 2012, S. 241f.). Dafür spricht zudem, dass Menschen scheinbar ungern vermenschlichende Zuschreibungen an Objekte eingestehen (Marquardt, 2017, S. 11f.). Entsprechend zeigten in nachfolgenden Untersuchungen indirekte Messinstrumente, die auf spontanen Antworten basierten, zumeist stärkere Indizien für Vermenschlichung als direkte Messinstrumente, die nur starken Anthropomorphismus abzubilden imstande sind (Fussell, Kiesler, Setlock & Yew, 2008).

Die CASA-These wurde seit ihrem Bestehen nicht dezidiert weiterentwickelt. Gambino, Fox und Ratan (2020) plädieren daher für eine Erweiterung und Überarbeitung der CASA-These, zumal sich in den vergangenen zwei Jahrzehnten die Denkweise der Menschen, die Interaktionsschnittstellen und -arten unterschiedlicher Technologien und die Technologien selbst verändert haben. Im Zusammenhang mit den Technologien sei nicht nur auf Fortschritte in der Robotik, sondern insbesondere auch im Bereich der Computer-Software, der virtuellen und erweiterten Realität sowie der natürlichsprachlichen Kommunikation verwiesen. Gleichsam entwickelten Menschen über die Zeit spezifische mentale Modelle und daraus generierte Verhaltensskripte für den Umgang mit modernen Medien, an deren Nutzung sie zunehmend gewöhnt sind. Möglicherweise lassen sich diese nun ebenso wie die Mensch-Mensch-Kommunikationsskripte unter gewissen Umständen gedankenlos anwenden und machen somit den Rückgriff auf soziale Interaktionsmuster obsolet (Gambino et al., 2020).

3.2.3.2 Theorie des dualen Anthropomorphismus

Einen kompatibilistischen Ansatz, der die *media equation* mit neueren Forschungen zum Anthropomorphismus in Einklang bringt, liefern Złotowski et al. (2018) mit ihrer Theorie des dualen Anthropomorphismus. Als Grundlage dafür diente den Autor:innen ein Experiment, in dem sich die Versuchsteilnehmenden in Abhängigkeit davon unterschiedlich verhielten, ob sie wussten, dass sie ihr Verhalten in Bezug auf eine Technologie rechtfertigen bzw. Fragen über deren Menschenähnlichkeit beantworten müssen. Złotowski et al. (2018) vermuten in einem *dual process*-Ansatz (vgl. z. B. Evans & Stanovich, 2013), dass zwei verschiedene Arten von Verarbeitungsprozessen an der Bildung der vermenschlichenden Zuschreibungen beteiligt sind. Die Klassifizierung eines mehrdeutigen Objekts als Mensch fassen sie als Ergebnis eines weitgehend automatisch und unbewusst ablaufenden Typ-1-Prozesses auf und nennen dieses Phänomen impliziten Anthropomorphismus. Diese Form des Anthropomorphismus sei entsprechend mittels indirekter Messmethoden feststellbar, die keine direkten Rückfragen stellen, sondern auf Schlüssen aus beobachtbaren Verhaltensweisen und Äußerungen basieren. Solche impliziten Anthropomorphisierungen können von bewussten und langsamer ablaufenden Typ-2-Überlegungen korrigiert werden, die allerdings einen gewissen kognitiven Aufwand erfordern und damit die nötige mentale Kapazität und Motivation zu dieser elaborierten Verarbeitung voraussetzen. Vermenschlichende Zuschreibungen, die auf Basis von Typ-2-Prozessen zustande kommen, werden als expliziter Anthropomorphismus bezeichnet. Derartige vermenschlichende Zuschreibungen seien über explizite Instrumente wie Selbstauskunftsskalen erhebbar (Złotowski et al., 2018).

Grundsätzlich können Typ-1- und Typ-2-Zuschreibungen koexistieren. Insbesondere wenn Rationalisierungen für vermenschlichende Zuschreibungen eingefordert werden, kann das Typ-2-Prozesse initiieren. Entsprechende Auskünfte können dann Ausdruck einer expliziten Vermenschlichung sein, wohingegen sich der implizite Anthropomorphismus im Verhalten widerspiegelt. Somit lässt sich u. a. der Befund aus der *media equation*-Forschung erklären, dass Personen sich so verhalten, als ob sie es mit einem sozialen Gegenüber zu tun hätten, allerdings gleichzeitig bei Nachfragen verneinen, dies zu tun. In diesen Fällen lägen Anzeichen für impliziten Anthropomorphismus vor, der allerdings durch die initiierten Typ-2-Verarbeitungsprozesse korrigiert würde, sodass sich kein expliziter Anthropomorphismus nachweisen ließe (Złotowski et al., 2018).

3.2.3.3 Drei-Faktoren-Theorie

Die sog. *Drei-Faktoren-Theorie* geht auf Arbeiten von Epley et al. (2007) und Epley et al. (2008) zurück. Mitunter wird diese Theorie in der Literatur auch mit dem Akronym der postulierten Faktoren als SEEK-Theorie bezeichnet (z. B. in Letheren, Kuhn, Lings & Pope, 2016). Die Motivation der Autor:innen bestand darin, die psychologischen Wirkmechanismen hinter dem Anthropomorphismus näher zu untersuchen, während sich ein

Großteil der Forschung damit beschäftigte, die Rechtfertigbarkeit von vermenschlichenden Zuschreibungen auf Basis der Charakteristika moderner Technologien zu bewerten (Epley et al., 2007; Waytz, Cacioppo & Epley, 2010).

Die Grundannahme der Theorie lautet, dass Anthropomorphismus durch das Zusammenspiel dreier Faktoren erzeugt wird. Dabei handelt es sich um einen kognitiven Faktor, der sich auf die Aktivierung menschenbezogenen Wissens (*elicited agent knowledge*) bezieht, und um zwei motivationale Faktoren, die in den entsprechenden Arbeiten empirisch getestet wurden, nämlich die Effektanzmotivation (*effectance motivation*) und die Sozialitätsmotivation (*sociality motivation*)². Zunächst legen Epley et al. (2007) dar, dass es für eine Person einen vergleichsweise geringen kognitiven Aufwand verursacht, auf Wissen über Menschen im Vergleich zu Wissen über Maschinen zurückzugreifen, um mittels dieses Wissens Schlussfolgerungen über das Verhalten eines Dritten anzustellen. Das führen sie darauf zurück, dass Menschen sich selbst am besten kennen (Nagel, 1974), über das Spiegelneuronensystem³ einen unmittelbaren phänomenologischen Zugang zu menschlichen Erlebenswelten haben und sich durch den von Geburt an vorhandenen Kontakt zu anderen Menschen ein reiches Repertoire an Erfahrungswerten ansammelt, um detaillierte und umfassende Modelle menschlicher Entitäten zu entwickeln (*exposure effect*). Schlussfolgerungen über das Verhalten von Dritten basieren demnach auf einem Prozess der egozentrischen Simulation und unterliegen durch die egozentrische Perspektivität entsprechenden Verzerrungen (*egocentric biases*). Ferner spezifizieren Epley et al. (2007) je Faktor dispositionale, situative, entwicklungspsychologische und kulturelle Aspekte, die die Wirkstärke des jeweiligen Faktors beeinflussen. In Bezug auf die Aktivierung menschenbezogenen Wissens gilt bspw. ein dispositionales hohes Bedürfnis nach Erkenntnis (*need for cognition*) als abschwächende Einflussvariable, da dieses Bedürfnis eine starke Affinität zu aufwändigen und reflexiven Denkprozessen impliziert. Für diese Prozesse ist keine schnelle Zugänglichkeit von Wissen über Menschen notwendig; vielmehr greifen sie auf abstraktes und schwerer zugängliches Wissen über maschinelle Entitäten zurück. Ferner ist als situative Einflussgröße die wahrgenommene Ähnlichkeit der Entität mit einem Menschen relevant sowie aus entwicklungspsychologischer Sicht die Verfügbarkeit von alternativen Theorien und Wissen über nicht-menschliche Entitäten, das sich die Person bereits angeeignet hat. Dass Erwachsene weniger zur Vermenschlichung neigen als Kinder, kann aus dieser Perspektive so verstanden werden, dass bessere Korrekturmechanismen für in gleichem Maße veranlagte Vermenschlichungsvorgänge bestehen (Epley et al., 2007; Epley, 2018). Als kultureller Aspekt ist ferner die Verbreitung anthropomorpher Erklärungsmuster sowie die Interaktionsmöglichkeiten mit modernen Technologien

² Die deutschen Übersetzungen sind angelehnt an Marquardt (2017, S. 37).

³ Das System der Spiegelneuronen gilt als wichtige neuronale Grundlage für das Verstehen von Handlungen und das Nachempfinden emotionaler Zustände fremder Lebewesen und für das Lernen durch Nachahmung; siehe hierzu bspw. Rizzolatti und Craighero (2004).

einzu beziehen. Epley et al. (2007, S. 870f.) gehen dabei davon aus, dass in hochindustrialisierten Ländern eine höhere Wahrscheinlichkeit besteht, nicht-anthropomorphe mentale Modelle von modernen Technologien zu entwickeln und diese verhältnismäßig leicht zu aktivieren.

Die isolierte Betrachtung dieses ersten Faktors lässt die Frage offen, wieso Personen überhaupt motiviert sein sollten, Vorhersagen über künftige Aktionen anderer Entitäten abzuleiten, zumal dieser Prozess eine vermeidbare kognitive Anstrengung erfordert. Eine Erklärung hierfür liefern die zwei motivationalen Komponenten, die sich einerseits auf das Bedürfnis, sich selbst als effektive Handelnde im Umgang mit der Umwelt zu erleben (*effectance motivation*), und andererseits auf das Bedürfnis nach sozialer Interaktion (*sociality motivation*) beziehen. Die Effektanzmotivation fällt umso höher aus, je stärker das Gefühl der Unsicherheit in Bezug auf eine fremde Entität ausgeprägt ist und je mehr deren Handlungen als unvorhersehbar eingeschätzt oder erlebt werden. Personen vermenschlichen Entitäten, die bspw. unvorhergesehene Fehler produzieren, nicht (primär) deswegen, weil sie Menschlichkeit mit Fehlerhaftigkeit und Unvorhersehbarkeit assoziieren, sondern aufgrund der Verunsicherung ob der zukünftigen Geschehnisse, die die Effektanzmotivation steigert (Waytz, Morewedge et al., 2010). Wenn die Technologie als menschlich repräsentiert wird, führt das zu einem Gefühl der Sicherheit im Umgang mit derselben. Als dispositionale Einflussvariable gilt daher der Wunsch nach Kontrolle (*desire for control*) und als situativer Faktor die (wahrgenommene) Vorhersagbarkeit (*apparent predictability*) der künftigen Aktionen der jeweiligen Entität. Entwicklungsdynamisch kann ein zunehmender Kompetenzerwerb (*attaining competence*) die Effektanzmotivation mindern, weil das Wissen um eigene Kompetenzen die Verunsicherung verringert. Der Grad, zu dem sich die Angehörigen einer Kultur durch Unsicherheit bedroht fühlen, stellt einen weiteren Einflussfaktor dar (Epley et al., 2007, S. 871ff.).

Vermenschlichung lässt sich vor dem Hintergrund der Effektanzmotivation funktional als Methode begreifen, um den zunehmenden technologischen Fortschritt und eine zunehmend technologisierte Umgebung zu bewältigen und sich in einer solchen Welt kompetent zu fühlen (Waytz, Morewedge et al., 2010). Allerdings stellt sich die Frage, wieso gerade die Anwendung eines menschenähnlichen Modells ein Gefühl der Effektanz vermittelt. Wenn gleich ein unvergleichlich hohes Maß an Vertrautheit mit Menschen kaum bestreitbar ist, bleiben menschliche Verhaltensweisen mitunter dennoch überraschend. Turkle (2010, S. 5) verweist explizit darauf, dass gerade jüngere Menschen die sozialen Interaktionen mit Mitmenschen als komplex und ermüdend empfinden (*fatigue*) und mitunter eine Präferenz zur als zunehmend vorhersehbar empfundenen Interaktion mit Technologie entwickeln. Sie befürchtet eine zunehmende Substitution echter zwischenmenschlicher Beziehungen durch quasi-soziale zwischen Menschen und Maschinen (Turkle, 2010, S. 5), in denen manche Menschen ihre Effektanzmotivation möglicherweise in höherem Ausmaße befriedigen zu können meinen. Das könnte zu einer Desozialisation der Gesellschaft führen, in der

Menschen perfekte und störungsfreie Beziehungen anstreben und durch überzogene Rationalisierung andere Personen dehumanisieren (de Graaf, 2016). Ob die Effektanzmotivation die Vermenschlichung befördert, hängt insofern stets davon ab, ob sich Personen eher im Umgang mit Maschinen oder mit Menschen als effektive Akteur:innen empfinden. Eine zunehmende Gewöhnung an eine technologisierte Welt könnte schließlich zu einem Vertrautheitsgefühl und einer geringeren Verunsicherung gegenüber modernen Technologien führen. Entsprechend deuten Waytz und Morewedge, et al. (2010) in einer Fußnote an, dass die Effektanzmotivation prinzipiell auch zur Aktivierung von mentalen Modellen von Maschinen führen kann.

Die Sozialitätsmotivation als weitere motivationale Komponente beschreibt das Bedürfnis, mit einem Gegenüber sozial zu interagieren, bspw. um ein Gefühl der Einsamkeit zu kompensieren. Entsprechend zeigte sich experimentell ein höheres Maß an Vermenschlichung von nicht-menschlichen Entitäten bei Versuchsteilnehmenden, die sich überdurchschnittlich einsam fühlten (Eyssel & Reich, 2013). Daher gilt chronische Einsamkeit (*chronic loneliness*) als dispositional begünstigender Faktor für die Sozialitätsmotivation, die durch das situative Empfinden sozialer Ausgrenztheit (*social disconnection*) noch verstärkt werden kann. Aus entwicklungspsychologischer Sicht wird Personen mit unsicher-ängstlichem Bindungsstil (*attachment style*) eine höhere Neigung zur Vermenschlichung zugeschrieben (Epley et al., 2007, S. 875ff.). Bemerkenswerterweise neigen einsame Menschen außerdem zur Dehumanisierung von Menschen, sodass sich in solchen Fällen eine grundsätzliche Zuwendung zu technischen Artefakten diagnostizieren lässt (Im Shin & Kim, 2018). Die Stärke der Sozialitätsmotivation ist ferner von der eher individualistischen oder kollektivistischen Ausrichtung einer Kultur abhängig, wobei Letztere den Wert von sozialen Beziehungen besonders akzentuiert (Epley et al., 2007, S. 877).

Zusammenfassend umschreiben die Effektanz- und die Sozialitätsmotivation, was Personen dazu veranlasst, die künftigen Aktionen von Interaktionspartnern vorhersagen und mit diesen sozial interagieren zu wollen. Dass gerade der Rückgriff auf menschenähnliche Repräsentanzen besonders nützlich ist, um diese Bedürfnisse zu befriedigen, lässt sich durch den kognitiven Vorteil der weniger aufwändigen Aktivierung menschenbezogenen Wissens erklären. Laut einer Studie von Im Shin und Kim (2018) stellt die Sozialitätsmotivation einen stärkeren Einflussfaktor auf den Anthropomorphismus dar als die leichtere Verfügbarkeit mentaler Modelle von Menschen. So neigten Informatik-Studierende, die über mehr Vorwissen im Umgang mit Technik verfügten, sich aber tendenziell einsamer fühlten als Studierende der Literaturwissenschaft, stärker zur Vermenschlichung (Im Shin & Kim, 2018).

Mit dieser Bestimmung und Gewichtung der begünstigenden Faktoren stellt sich die Frage nach der Universalität des Anthropomorphismus. Die Tendenz zur Vermenschlichung scheint zwar in allen Menschen angelegt zu sein, die Intensität allerdings personen-

individuell unterschiedlich auszufallen. Dabei gelten die Persönlichkeit, die Denkweise, das Alter, der Beziehungsstatus und die emotionale Bindung bzw. Bindungsfähigkeit zu Tieren als Einflussfaktoren bzw. konfundierende Variablen in Bezug auf den Anthropomorphismus (Letheren et al., 2016). Mit der IDAQ-Skala (*individual differences in anthropomorphism*) steht ein Messinstrument zur Verfügung, um diese Veranlagung personenindividuell zu erheben. Personen mit einer hohen dispositionalen Neigung zur Vermenschlichung vertrauen dabei auch eher auf Technologien (Waytz, Cacioppo & Epley, 2010).

3.2.3.4 Modell der Frame-Selektion

Auf Basis des soziologisch orientierten *Modells der Frame-Selektion* (MdFS) nach Esser (2002) und Esser (2005) lässt sich Anthropomorphismus konzeptionell auch als Selektionsproblem fassen, bei dem zwischen der Anwendung eines menschen- und eines maschinenähnlichen mentalen Modells zu entscheiden ist. Innerhalb des MdFS werden diese als Frames bezeichnet, die die Situation für die jeweiligen Agierenden definieren. Damit ist sprachlich bereits eine Nähe zu Theorien des sprachlichen Framings (vgl. Kapitel 3.3) gegeben. Passend zu den Erläuterungen in Kapitel 3.1, lässt sich auf Basis der Beschreibungen von Frames im Rahmen des MdFS vermuten, dass diese erheblichen Überschneidungen zum Konzept der mentalen Modelle aufweisen und auch als solche verstanden werden können. In der folgenden Beschreibung wird zwar der Begriff des Frames theoriekongruent beibehalten, allerdings unter der Prämisse, dass damit eine Repräsentationsstruktur beschrieben wird, die mentalen Modellen gleicht und in der vorliegenden Dissertation unter diesem Begriff gefasst wird.

Da die beschriebenen Frames mit gewissen erwarteten Hinweisreizen verknüpft sind, werden sie bei deren Wahrnehmung spontan aktiviert und stehen damit intuitiv als Deutungsmechanismen für die erlebte Situation zur Verfügung, analog zu einem situativ aktivierten mentalen Modell. Ferner ist bei der Wahrnehmung solcher Mensch-Technik-Interaktionssituationen stets ein zweites Selektionsproblem zu lösen, nämlich die Auswahl eines Informationsverarbeitungsmodus, der als Heuristik angewendet wird, um die vorliegende Situation kognitiv zu erfassen und zu bewältigen. Hierzu stehen entweder der automatisch-spontane (*as*) oder der reflektiert-kalkulierende Modus (*rc*) zur Verfügung, die sich jeweils durch ihre Schnelligkeit in der Ausführung und die Elaboriertheit der Informationsverarbeitung unterscheiden (Esser, 2005). Diese Unterscheidung weist Parallelen zur Theorie des dualen Anthropomorphismus auf, die ebenfalls verschiedene Verarbeitungsmodi ins Zentrum stellt. Im *as*-Modus ist sowohl die Verarbeitung kognitiver als auch emotionaler Reaktionen vorgesehen.

Letztlich bestimmen mehrere Faktoren, welche Selektionen vorgenommen werden. Diese Faktoren lassen sich mitunter den Faktoren aus der Drei-Faktoren-Theorie zuordnen. So wird die Selektion eines Frames u. a. durch dessen wahrgenommene Passung zur vorgefun-

denen Situation (*match*) bestimmt, die sich wiederum aus der kognitiven Zugänglichkeit der Frames (vgl. *elicited agent knowledge*), dem Vorhandensein prototypischer Symbole innerhalb der Situation, bspw. durch ein anthropomorphes Design, und aus der Abwesenheit von „Störungen der gewohnten Umstände“ (Esser, 2005, S. 7) ergibt. Diese Passung ist für die jeweilige Person nicht bewusst kognitiv beeinflussbar, sondern „, passiert‘ oder wird ‚erlebt‘“ (Esser, 2005, S. 8). Die erwartete Nützlichkeit eines bestimmten Frames zur Bewältigung der Situation lässt sich nun mit dessen Passung multiplizieren, um den Erwartungsnutzen dieses Frames zu bestimmen. Dabei lässt sich die Nützlichkeit als Erfüllungsgrad der motivationalen Faktoren aus der Drei-Faktoren-Theorie konzeptualisieren. Wenn ein alternativer Frame als nützlicher bewertet wird und eine ausreichend hohe Passung zur Situation aufweist, kann dies einen Anreiz für ein *reframing* der Situation darstellen. Umgekehrt wird ein gegenwärtiger Frame als salient bezeichnet, wenn sich bei seiner Anwendung keine hinreichenden Motivationen für ein *reframing* ergeben (Esser, 2005).

Für die Auswahl des Verarbeitungsmodus gilt, dass in der Regel der saliente Frame automatisch zur Anwendung kommt. Nur im Falle einer hinreichend starken Reflexionsmotivation findet ein Wechsel in die *rc*-Verarbeitung statt, im Rahmen derer unter Inkaufnahme eines gewissen kognitiven Aufwands eine alternative Situationsdefinition gesucht wird. Die Stärke der Reflexionsmotivation hängt von mehreren Faktoren ab. Hierzu zählt die geschätzte Wahrscheinlichkeit, einen passenderen Alternativframe zu identifizieren, dessen Nutzen für die Bewältigung der vorliegenden Situation die Reflexionskosten im Sinne eines kognitiven Verarbeitungsaufwands mutmaßlich übersteigt (Esser, 2005). Stark saliente Frames oder komplexe Interpretationslagen, die hohe Reflexionskosten implizieren, steigern die Wahrscheinlichkeit einer *as*-Verarbeitung, die als Standardmodus betrachtet werden kann. Insofern bietet die MdFS eine stark formalisierte Konzeptualisierung von Anthropomorphismus, die in hohem Maße an die Theorie der mentalen Modelle und des dualen Anthropomorphismus sowie an die Drei-Faktoren-Theorie anschlussfähig bzw. damit kompatibel ist (Marquardt, 2017, S. 31).

3.2.4 Anthropomorphisierung von Robotern

3.2.4.1 Konventionelle Roboter und Cobots als Vermenschlichungsobjekte

Typischerweise beziehen sich vermenschlichende Zuschreibungen v. a. auf übernatürliche Wesen, die Natur und damit zusammenhängende Phänomene sowie (Haus-)Tiere, aber auch auf komplexe maschinelle Systeme wie bspw. (vermeintlich) intelligente Software-Agenten oder Roboter (Marquardt, 2017, S. 8; Złotowski et al., 2018). In einer Medienanalyse auf Basis von Veröffentlichungen der *International Conference on Social Robotics* belegt Zijlstra (2017), dass sogar Wissenschaftler:innen die von ihnen entwickelten Roboter häufig mit vermenschlichenden Attributen beschreiben. Wie bereits in Kapitel 2.2.4 im Zusammenhang mit hybriden Robotern angedeutet, verfügen Roboter über Merkmale, die

Vermenschlichung begünstigen, nämlich bspw. eine Verkörperung, ein oftmals anthropomorphes Aussehen sowie mitunter unvorhersehbare menschenähnliche Bewegungen (Bartneck et al., 2005; Ferrari & Eyssel, 2016; Marquardt, 2017, S. 30; Złotowski et al., 2014). Daher eignen sich Roboter letztlich „*par excellence* als Objekte der Anthropomorphisierung“ (Marquardt, 2017, S. 4), wenngleich das Vorhandensein der zuvor aufgezählten Merkmale weder eine notwendige noch eine hinreichende Bedingung für den psychischen Prozess der Vermenschlichung darstellt. In diesem Zusammenhang lassen sich Roboter auch als evokative Objekte interpretieren (Turkle, 2007), die gar zu „Gefährten des Gefühlslebens“ (Marquardt, 2017, S. 5) avancieren und für den Umgang mit Tieren und Menschen reservierte Gefühlskategorien hervorrufen können. Letzteres bestätigte sich auch in neurowissenschaftlichen Studien unter Zuhilfenahme der funktionellen Magnetresonanztomographie (fMRT), in denen die Versuchsteilnehmenden nicht nur bei der Beobachtung einer gewaltsamen Handlung gegenüber einem Menschen, sondern auch gegenüber einem Roboter emotionalen Stress zeigten (Rosenthal-von der Pütten et al., 2014). Diese affektive Stressreaktion trotz der kognitiven Einsicht, dass ein Roboter nicht leidensfähig ist, lässt sich mit dem Modell des dualen Anthropomorphismus (vgl. Kapitel 3.2.3.2) in Einklang bringen. Manche Forschende erachten es daher gar für sinnvoll, neue ontologische Kategorien für gewisse Roboter zu etablieren, die der Ambiguität aus *de facto*-Unterschiedlichkeit und wahrgenommener Ähnlichkeit mit Menschen Rechnung tragen (Bergamasco & Herr, 2016, S. 1902; Ferrari & Eyssel, 2016) und Zustände zwischen *unbelebt* und *lebendig* abbilden (Melson et al., 2005). In einem solchen Zwischenstadium verorteten 9- bis 15-jährige Kinder im Rahmen einer Interaktionsstudie einen beteiligten Roboter (Kahn et al., 2012).

Im Kontext der vorliegenden Dissertation ist von besonderem Interesse, inwiefern die Merkmale von Cobots Anthropomorphismus begünstigen. Remmers (2020) weist darauf hin, dass abgesehen von den o. g. typischen und augenfälligen Merkmalen humanoider Roboter bereits die interaktive Ausgestaltung der Anwendungsszenarien, an den Menschen angepasste Roboterarmbewegungen oder schlicht das Repertoire verschiedener Tätigkeiten als subtile Trigger der Vermenschlichung wirken können. Tabelle 7 bietet eine Übersicht darüber, inwiefern sich die entsprechenden Merkmale bei Cobots wiederfinden lassen, und liefert damit einen Beleg dafür, dass sich Cobots noch besser als Vermenschlichungsobjekte eignen als konventionelle Roboter.

Tabelle 7: Design-Merkmale von Robotern, die Vermenschlichung begünstigen, in Bezug zu (industriellen) Cobots.

Merkmale	Bezug zu (industriellen) Cobots
Physische Verkörperung	Bei Cobots wie bei jedem Roboter vorhanden.
Häufig anthropomorphes Aussehen	<ul style="list-style-type: none"> • Teilweise gezielt anthropomorphe Merkmale wie Display mit Augenpaaren (<i>Baxter</i>, <i>Sawyer</i>). • Teilweise morphologisch bedingt menschenähnliche Merkmale wie Torso mit zwei Armen (<i>yumi</i>).
Menschenähnliche Bewegungen	<ul style="list-style-type: none"> • Vergleichsweise langsame und gleichmäßige Bewegungen in direkter Umgebung des Menschen. • Ggf. an den Menschen angepasste Trajektorien, z. B. Bewegungen auf den Menschen zu, um Bauteile gemeinsam bearbeiten zu können.
Interaktive menschenähnliche Tätigkeiten	<ul style="list-style-type: none"> • Tätigkeiten, die menschlichen Verhaltensweisen stark ähneln, wie z. B. das Greifen, Ausrichten und Absetzen kleiner Gegenstände. • Reichhaltiges Repertoire solcher Tätigkeiten. • Hohes Maß an Interaktivität, u. a. durch die Möglichkeit direkter Berührungen.

Im industriellen Umfeld, in dem Cobots im Regelfall eingesetzt werden, wirken vermenschlichende Zuschreibungen „zunächst paradox“ (Onnasch et al., 2019, S. 33), da Roboter in diesem Bereich im Gegensatz zu sozialen Robotern primär für den funktionalen Einsatz optimiert sind und „bisher deutlichen Werkzeugcharakter“ (Onnasch et al., 2019, S. 33) haben. Formulierungen aus dem Bereich der MRK wie „hybride[] Teams‘ oder ‚Roboter[] als Kollegen“ (Onnasch et al., 2019, S. 33) verdeutlichen die Tendenz zur sprachlichen Vermenschlichung von Robotern im Industrieumfeld und durchbrechen die bisher gängige Konzeption als reines Werkzeug. Die Notwendigkeit der Interaktion fördert die Zuschreibung von bestimmten Zielen und Absichten an den Roboter, um Handlungen aufeinander abstimmen zu können (Onnasch et al., 2019, S. 33). In der industriellen Praxis gelten vermenschlichende Zuschreibungen als Beleg für eine hohe Akzeptanz, wie eine Aussage aus einem Interview mit einem Unternehmensvertreter belegt: „Wenn die Roboter (Kose-)Namen bekommen, ist die Technologie bei den Mitarbeitern angekommen“ (Bender et al., 2016, S. 28). Wenngleich diese vermenschlichenden Zuschreibungen kontraintuitiv erscheinen, lässt sich ihr Vorkommen mittels der Drei-Faktoren-Theorie im Kontext der industriellen MRI unter Voraussetzung einiger plausibler Annahmen gut erklären, wie Tabelle 8 zeigt. In Hinblick auf die Sozialitätsmotivation ist dabei insbesondere zu

bedenken, dass Ängste vor sozialer Isolation bestehen, die die Arbeitszufriedenheit bedrohen können (Elprama et al., 2017; Heßler, 2015; Müller-Abdelrazeq, 2020).

Tabelle 8: Faktoren für Vermenschlichung laut Drei-Faktoren-Theorie und deren vermutete Relevanz im Kontext der industriellen MRI.

Merkmale	Bezug zum Kontext der industriellen MRI
Aktivierung menschenbezogenen Wissens	<ul style="list-style-type: none"> • Aufgrund mangelnder Erfahrung und Neuheit der Technologie bestehen wenige konkrete mentale Modelle von Cobots. • Die im industriellen Kontext tendenziell geringe dispositionale Neigung zur kognitiv aufwändigen Verarbeitung begünstigt die Aktivierung einfach zugänglicher Repräsentationen von Menschen. • Grundprinzip der Minimierung des kognitiven Aufwands anwendbar.
Effektanzmotivation	<ul style="list-style-type: none"> • Kooperatives Arbeiten erfordert hohe Vorhersagegenauigkeit der Aktionen eines Roboters. • Hohe Vorhersagegenauigkeit der Aktionen des Cobots ist notwendig, um Verletzungen vorzubeugen. • Einschätzung von Funktionsweise und Leistungsfähigkeit ist notwendig, um Wahrscheinlichkeit technologischer Ersetzung abschätzen zu können.
Sozialitätsmotivation	<ul style="list-style-type: none"> • Soziale Interaktion stellt einen wichtigen Faktor der Arbeitszufriedenheit dar. • Im Rahmen der Technisierung der Arbeit ist die Sorge vor sozialer Isolation verbreitet.

3.2.4.2 Drei-Phasen-Modell der Vermenschlichung

Die bisher vorgestellten theoretischen Ansätze erklären Anthropomorphismus als situatives Phänomen. Diese Ansätze lassen allerdings eine prozessorientierte Sicht auf Vermenschlichungsvorgänge vermissen, die im Zeitverlauf dynamischen Veränderungen unterliegen. Diese Lücke versuchten Lemaignan, Fink & Dillenbourg (2014) zu schließen, indem sie ein Modell entwarfen, das Anthropomorphismus als ein sich in drei Phasen entwickelndes, „dynamisches, nicht-monotones und kontext-abhängiges Phänomen“ (*dynamic, non-monotonic and context-dependent process*) (Lemaignan, Fink & Dillenbourg, 2014, S. 226) beschreibt.

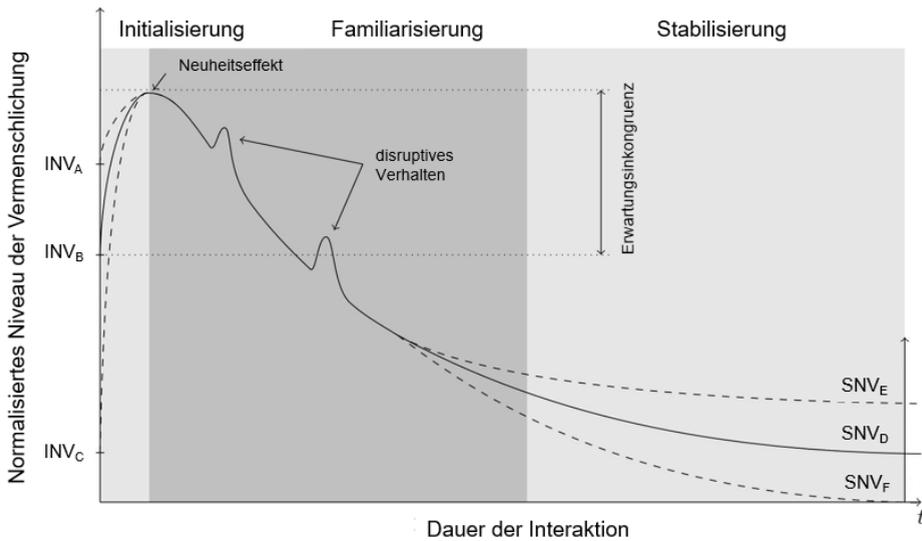


Abbildung 3.1: Dynamischer Verlauf der Intensität von Vermenschlichung in Abhängigkeit von der Interaktionsdauer. Übersetzt und leicht modifiziert nach Lemaignan, Fink, Dillenbourg und Braboszcz (2014, S. 2).

Lemaignan, Fink & Dillenbourg (2014) entwickelten ihr Modell im Gegensatz zu den vorherigen Ansätzen anhand einer typischen MRI, bezogen sich also direkt auf Roboter. Die MRI-Situation unterteilten sie in die drei Phasen der Initialisierung, Familiarisierung und Stabilisierung, wie in Abbildung 3.1 ersichtlich. Bereits vor dem Beginn einer MRI besteht zum Zeitpunkt des ersten (realen oder imaginierten) Kontakts mit dem Roboter ein initiales Vermenschlichungsniveau (INV ; *initial capital of anthropomorphism*). Die Autor:innen beschreiben dieses Niveau als situationsspezifisch vorhandenes Anthropomorphisierungspotenzial einer Person in Bezug auf einen bestimmten Roboter (Lemaignan, Fink, Dillenbourg & Braboszcz, 2014, S. 1f.). Ausgehend von diesem Grundniveau erreicht die Vermenschlichung zu Beginn einer MRI im Laufe einer Initialisierungsphase, die typischerweise wenige Minuten bis einige Stunden andauert, aufgrund eines Neuheitseffekts (*novelty effect*) einen Höhepunkt. Daraufhin ist in der einige Tage andauernden Familiarisierungsphase ein stetiges Absinken zu erwarten. Dies begründen die Autor:innen damit, dass die Aktionen des Roboters zunehmend verständlicher und vorhersagbarer werden und sich damit der psychologische Anreiz reduziert, ein menschenähnliches mentales Modell zur Vorhersage in Anwendung zu bringen. Unvorhergesehene Aktionen des Roboters (*disruptive behaviors*) können dabei laut dem Drei-Phasen-Modell zu jedem Zeitpunkt einen erneuten sprunghaften Anstieg der Vermenschlichung hervorrufen. Zwar liefern Lemaignan, Fink, Dillenbourg und Braboszcz (2014) keine Begründung für den

postulierten Anstieg des Vermenschlichungsniveaus durch die unvorhergesehenen Aktionen, allerdings lässt sich dieser Effekt auf Basis empirischer Studien damit begründen, dass die Verunsicherung aufgrund der unerwarteten Aktionen die Effektanzmotivation und den damit verbundenen Anreiz zur Vermenschlichung verstärkt (Waytz, Morewedge et al., 2010).

Im Verlauf einer länger währenden Stabilisierungsphase sinkt das Vermenschlichungsniveau weiter ab, bis ein individuell unterschiedliches und vom Eingangsniveau unabhängiges stabilisiertes Niveau der Vermenschlichung (SNV; *stabilized level of anthropomorphism*) erreicht ist. Während zu Beginn der MRI das Erscheinungsbild des Roboters (*appearance*) einen stärkeren Einfluss ausübt, rücken während der MRI zunehmend die Aktionen des Roboters (*behavior*) und insbesondere deren Vorhersehbarkeit in den Fokus (Lemaignan, Fink, Dillenbourg & Braboszcz, 2014). Entsprechend zeigte sich in einer empirischen Studie, dass initial unterschiedlich hohe Vermenschlichungsniveaus in Bezug auf unterschiedlich gestaltete Roboter während einer realen Interaktion auf ein gemeinsames Niveau konvergierten (Abubshait & Wiese, 2017). Der in der Abbildung verzeichnete Effekt der Erwartungsinkongruenz (*expectation mismatch*) wird von den Autor:innen nicht näher erläutert und lässt sich daher nur spekulativ als Differenz zwischen dem Eingangswert der Vermenschlichung und dem durch den Neuheitseffekt erzeugten Höhepunkt interpretieren, der zu überhöhten Erwartungen führt, die sich in der späteren Beobachtung nicht bewahrheiten.

Den skizzierten Verlauf begründen Lemaignan, Fink & Dillenbourg (2014) mit der Existenz dreier unterschiedlicher kognitiver Phasen, die sich allerdings nicht vollständig mit den vorgenannten zeitlichen Phasen decken. Die Initialisierungsphase ist von einer subkognitiven Vermenschlichung gekennzeichnet, die instinktiv stattfindet. In einer zweiten Phase findet eine Projektion existierender mentaler Modelle auf den vorgefundenen Roboter statt, wobei hier die Anwendung anthropomorpher Modelle die starke Vermenschlichung begründet (vgl. Drei-Faktoren-Theorie in Kapitel 3.2.3.3). In der Familiarisierungsphase wird das mentale Modell stärker an die realen Erfahrungen angepasst, sodass sich ein spezifisch auf den Roboter passendes Modell ausbildet. In einer dritten kognitiven Phase werden nur noch marginale Verfeinerungen des mentalen Modells unter Berücksichtigung des Interaktionskontextes und der -ziele vorgenommen. Sobald sich das mentale Modell hinreichend gefestigt hat und durch weitere Erfahrungen nicht mehr signifikant verändert, ist das stabilisierte Vermenschlichungsniveau erreicht. Die Anzahl der in diesem Modell repräsentierten anthropomorphen Merkmale bestimmt die Intensität der Vermenschlichung (Lemaignan, Fink, Dillenbourg & Braboszcz, 2014).

Marquardt (2017, S. 24) diagnostiziert eine hohe Kompatibilität zwischen dem Drei-Phasen-Modell und der Drei-Faktoren-Theorie, auf die Lemaignan, Fink & Dillenbourg (2014) allerdings nicht explizit Bezug nehmen. Da im Laufe der Familiarisierungsphase ein

mentales Modell vom Roboter als Erklärungsheuristik für dessen Aktionen zur Verfügung steht, lässt sich die Vermenschlichung zu diesem Zeitpunkt nicht mehr mittels der Effektmotivation und der einfacheren Aktivierbarkeit des Wissens über Menschen begründen. Daher liegt die Vermutung nahe, dass zwar die Varianz des Eingangsniveaus der Vermenschlichung durch diese Faktoren begründbar ist, die Abweichungen in den stabilisierten Niveaus der Vermenschlichung allerdings primär die individuell unterschiedlich ausgeprägte Sozialitätsmotivation abbildet (Marquardt, 2017, S. 23). Insbesondere der postulierte Neuigkeitseffekt ist aus Forschungssicht relevant, da in Laborsituationen häufig nur kurze Erstkontakte betrachtet werden, die demzufolge nur das Maximum des Vermenschlichungsniveaus messen (Marquardt, 2017, S. 25). Vor diesem Hintergrund scheint es gewinnbringend, die zeitliche Dynamik von Vermenschlichung mitzudenken, wofür das Drei-Phasen-Modell einen ersten Ansatz liefert. Einschränkend ist allerdings hinzuzufügen, dass das Modell eher auf anekdotischen Evidenzen und theoretischen Überlegungen denn auf solider Empirie basiert und von den Autor:innen selbst als „spekulativ“ (*speculative*) (Lemaignan, Fink, Dillenbourg & Braboszcz, 2014, S. 3) bezeichnet wurde. Das Modell wurde in der Folge nur in geringem Umfang rezipiert und weiterentwickelt, sodass bspw. mögliche Faktoren, die die Anthropomorphisierung mit fortschreitender Interaktionsdauer verstärken könnten, wie etwa eine stärkere soziale Bindung an eine Technologie, weder thematisiert noch einbezogen werden.

3.2.4.3 Typische Folgen der Vermenschlichung

Vermenschlichende Zuschreibungen haben verhaltensbezogene Konsequenzen zur Folge, die sich u. a. in einem verstärkten Sozialverhalten gegenüber vermenschlichten Robotern zeigen können. Neuere Studien im MRI-Bereich rekurren in diesem Zusammenhang auf die grundsätzlichen Annahmen aus der *media equation*-Theorie (u. a. Brandstetter, 2017; Horstmann et al., 2018), zu denen sich inzwischen eine große Literaturbasis und empirische Fundierung entwickelt hat (Horstmann et al., 2018; Rosenthal-von der Pütten et al., 2014). Broadbent (2017, S. 640f.) liefert hierzu eine Sammlung von Einzelstudien (vgl. Tabelle 9), die verschiedene Arten des Sozialverhaltens gegenüber Robotern dokumentieren.

Tabelle 9: Experimentelle Beispiele für Sozialverhalten gegenüber Robotern.

Beobachtbares Sozialverhalten	Beschreibung	Quelle
Anwendung von Vorurteilen bzgl. (wahrgenommener) Nationalität	Deutsche Versuchspersonen bewerten einen in Deutschland hergestellten und mit deutschem Namen versehenen Roboter als	Eyssel und Kuchenbrandt (2012)

Beobachtbares Sozialverhalten	Beschreibung	Quelle
	warmherziger und leistungsfähiger.	
Anwendung von Vorurteilen bzgl. (wahrgenommenem) Geschlecht	Ein weiblich gestalteter Roboter wird als geeigneter für bestimmte Tätigkeiten eingeschätzt, die überwiegend von Frauen verrichtet werden (z. B. Kindererziehung).	Eysssel und Hegel (2012)
	Ein Roboter mit weiblicher Stimme erhält höhere Geldspenden als ein Roboter mit männlicher Stimme.	Siegel, Breazeal & Norton (2009)
Höflichkeit	Versuchspersonen verhalten sich höflich gegenüber einem <i>Nao</i> -Roboter und versuchen ihm zu helfen.	Rehm & Krogsgager (2013)
	Versuchspersonen begrüßen einen Roboter.	M. K. Lee, Kiesler & Forlizzi (2010)
Reziprozität	Ein Roboter wird positiver bewertet, wenn er den Versuchspersonen einen Gefallen tut, nachdem sie ihm geholfen haben.	Lammer, Huber, Weiss & Vincze (2014)
	Versuchspersonen zeigen reziprokes Verhalten im Gefangenendilemma, vergleichbar mit dem Verhalten gegenüber Menschen.	Sandoval, Brandstetter, Obaid und Bartneck (2016)
Befolgung sozialer Normen	Versuchspersonen verhalten sich ehrlicher und folgen weniger dem Anreiz zu betrügen, wenn ein Roboter anwesend ist, als wenn sie allein sind.	Hoffman et al. (2015)

Grundsätzlich wird davon ausgegangen, dass sich die zuvor beschriebene CASA-These auch auf Roboter anwenden lässt. Allerdings liegen auch einige kontroverse empirische Befunde vor. So zeigten Versuchsteilnehmende unterschiedliche Erwartungen an eine bevorstehende Interaktion mit Robotern im Gegensatz zu Menschen (A. Edwards et al., 2019). Ferner lieferte die linguistische Analyse der sprachlichen Interaktionen zwischen verschiedenen Versuchsteilnehmenden und einem Roboter in einer Studie Hinweise

darauf, dass die beteiligten Personen interindividuell unterschiedlich stark sozial auf den Roboter reagierten (Fischer, 2011). Im gleichen Experiment fiel auf, dass viele Versuchspersonen lachten, bevor sie dem Roboter antworteten, was als Zeichen dafür gewertet werden kann, dass das Interaktionsverhalten als ein Akt des gemeinsamen Vortäuschens (*joint pretense*) zu verstehen ist. In einem weiteren Experiment zeigten Versuchspersonen ein unterschiedliches Kommunikationsverhalten in Abhängigkeit davon, ob sie mit einem Chatbot oder einem Menschen kommunizierten (Mou & Xu, 2017). Außerdem ergab eine Replikation des bekannten Milgram-Experiments (Milgram, 1963), dass die Versuchsteilnehmenden einen Roboter signifikant höheren Elektroschocks aussetzten als einen Menschen (Bartneck et al., 2005). Dieses unterschiedliche Verhalten gegenüber Menschen und Robotern bewerten die Autor:innen als Einschränkung des Anwendungsgebiets der *media equation*. Nichtsdestotrotz zeigten die beteiligten Versuchsteilnehmenden Mitgefühl gegenüber dem Roboter, der den Elektroschocks ausgesetzt wurde, obgleich sie sich bewusst waren, dass dieser nicht leidensfähig ist.

Gemeinhin wird Anthropomorphismus positiv als hilfreicher Mechanismus zur Etablierung einer gelingenden Mensch-Roboter-Interaktion bewertet und mit wünschenswerten Folgen in Verbindung gebracht (Zawieska et al., 2012), wie z. B. einem höheren Vertrauen (Waytz, Cacioppo & Epley, 2010) oder einer höheren Akzeptanz und Fehlertoleranz (Darling, 2017, S. 4; Kory Westlund et al., 2016). Gleichwohl birgt Anthropomorphismus die Gefahr überzogener Erwartungen und kognitiver Dissonanzen, wenn einzelne Merkmale eines Roboters (wie Aussehen und Bewegung) unterschiedliche mentale Modelle aktivieren (Wiese et al., 2017). Wenn Roboter zu menschenähnlich wahrgenommen werden, kann das zu unrealistischen Erwartungen führen, deren Nichterfüllung Frustrationen zur Folge hat (Frennert, Efring & Östlund, 2017).

Das sog. *Schema der perfekten Automation* (*perfect automation schema*; PAS) stellt eine relevante Heuristik zur Erklärung von Vermenschlichungseffekten dar, die insbesondere eine höhere Fehlertoleranz gegenüber vermenschlichten Robotern zu erklären vermag. Aus der Forschung zu Vertrauen in Automationssysteme ist ein sog. *Positivitätsbias* bekannt (Dzindolet, Pierce, Beck & Dawe, 2002). Dieses bezieht sich auf die Tendenz, solchen Systemen mehr Vertrauen zu schenken als Menschen, da diesen Systemen im Gegensatz zu Menschen eine perfekte Funktionsfähigkeit zugeschrieben wird. Technische Systeme aktivieren Assoziationen der Perfektion, wohingegen Menschen als fehleranfällig und in dieser Hinsicht defizitär eingeschätzt werden (Madhavan & Wiegmann, 2007). Entsprechend bestehen häufig überhöhte Erwartungen an und ein übersteigertes initiales Vertrauen in Technologien, woraus eine erhöhte Sensitivität gegenüber Fehlern von Automatisierungssystemen resultiert, da diese Fehler als widersprüchlich zu dem entsprechenden mentalen Modell empfunden werden (W. Kim, Kim, Lyons & Nam, 2020; Lyons & Guznov, 2019; Madhavan & Wiegmann, 2007; Merritt, Unnerstall, Lee & Huber, 2015). Dieses Bias wird in vielen aktuellen Publikationen thematisiert, wenngleich auch abweichende

Forschungsergebnisse vorliegen, die den Bedarf nach weiterer Forschung unterstreichen (Madhavan & Wiegmann, 2007, S. 286). Einige Autor:innen gehen davon aus, dass Menschen dieses Schema der perfekten Automation auch auf Roboter anwenden, wenngleich diese eine tendenziell menschenähnlichere Erscheinung mit sich bringen als klassische Automationsysteme (Alarcon, Capiola & Pfahler, 2021, S. 173; Lyons & Guznov, 2019). Eine stärkere Vermenschlichung solcher Systeme könnte das Positivitätsbias in Hinblick auf diese Robotern vermindern, sie also weniger perfekt und fehleranfälliger erscheinen lassen (Madhavan & Wiegmann, 2007). Aus der Annahme einer höheren Fehleranfälligkeit ergäbe sich dann eine höhere Fehlertoleranz im Verlauf einer Interaktion.

Tabelle 10 und Tabelle 11 stellen die typischerweise genannten Auswirkungen von Vermenschlichung und deren übliche Bewertung als positive oder negative Folgen dar. Einschränkung sei hierbei auf die Unterkomplexität dieser Einteilung verwiesen, da die Folgen kontextabhängig unterschiedlich zu bewerten sind. So können etwa eine starke emotionale Bindung oder ein hohes Vertrauensniveau nachteilig und ethisch bedenklich sein, insbesondere wenn die Technologie ein solches Vertrauen nicht rechtfertigt und Bindungsbedürfnisse nur scheinbar befriedigen kann. Turkle (2011, S. 6) weist in diesem Zusammenhang darauf hin, dass es in Mensch-Roboter-Relationen auch für gute Gefühle schlechte Gründe geben kann, z. B. wenn das Wohlgefühl aus der bloßen Überzeugung entspringt, das Gegenüber kontrollieren zu können.

Tabelle 10: Mögliche (vermeintlich) positive Auswirkungen von Anthropomorphismus.

Vermeintlich positive Auswirkungen	Quelle
Höheres Vertrauen (u. a. in die Leistungsfähigkeit bei verschiedenen Aufgaben) und höhere Akzeptanz	Waytz et al. (2014); Kory Westlund et al. (2016)
Stärkeres Wohlgefühl während der Interaktion	Abubshait und Wiese (2017)
Höhere emotionale Bindung	Groom et al. (2009)
Höhere Fehlertoleranz	Darling (2017); vgl. Schema der perfekten Automation (Dzindolet et al., 2002)
Flüssigere Interaktion durch intuitive Benutzerschnittstelle	Schellen, Pérez-Osorio und Wykowska (2018); Stadler et al. (2013); Hoffman (2019)

Tabelle 11: Mögliche (vermeintlich) negative Auswirkungen von Anthropomorphismus.

Vermeintlich negative Auswirkungen	Quelle
Irreführende Erwartungen und entsprechende Frustrationen durch Verletzung der Erwartungskonformität	Charalambous et al. (2016a)
Unheimliche Erscheinung aufgrund des Effekts des unheimlichen Tals (vgl. Kapitel 3.2.4.4)	Mori (1970)
Weniger stark wahrgenommener funktionaler Wert	Onnasch und Roesler (2019); Roesler und Onnasch (2020a)
Kognitive Dissonanz durch konfligierende anthropomorphe und nicht-anthropomorphe Merkmale	Wiese et al. (2017)
Geringeres Situationsbewusstsein	Hinds et al. (2004); Gombolay, Bair, Huang und Shah (2017)
Erwartung von Fehlern	Vgl. Schema der perfekten Automation (Dzindolet et al., 2002)

Die Auswirkungen von Vermenschlichung sollten demnach nicht pauschal, sondern nur kontextbezogen beurteilt werden. Hierbei ist in erster Linie der Zusammenhang zwischen vermenschlichenden Zuschreibungen und daraus entstehenden Erwartungen entscheidend. In ihrer *Passungshypothese (matching hypothesis)* unterstreichen Goetz, Kiesler und Powers (2012), dass der Erfolg einer MRI von der Übereinstimmung zwischen Design und Aktionen eines Roboters mit den Erwartungen des Nutzers vor dem Hintergrund des Aufgabenkontexts abhängt (Marquardt, 2017, S. 10). In Abhängigkeit von der jeweiligen Aufgabe und dem relativen Hierarchiestatus des Roboters (als Weisungsgeber:in oder -empfänger:in) können dabei entweder menschenähnliche oder maschinelle Repräsentationen von Robotern eher geeignet erscheinen (Hinds et al., 2004). Diese Erkenntnis weist bereits darauf hin, dass bei der Abschätzung der Auswirkungen von Anthropomorphisierung stets das Ausmaß der Vermenschlichung sowie etwaige Interaktionseffekte mit weiteren Faktoren zu berücksichtigen sind. Entsprechend fanden Yogeewaran et al. (2016) experimentell einen Interaktionseffekt zwischen der wahrgenommenen Menschenähnlichkeit und der Sorge, von einem Roboter übertroffen zu werden, der dazu führen kann, dass die Effekte von Vermenschlichung ab einer gewissen Schwelle ins Negative umschlagen. Dieser Schwellenwert lässt sich allerdings nicht pauschal bestimmen, sondern ist seinerseits wiederum vom subjektiven Empfinden und dem Aufgabenkontext abhängig (Fink, 2012, S. 201). Daher empfiehlt es sich, anthropomorphes Design nicht als Selbstzweck zu betrachten, sondern allenfalls zielgerichtet und in angemessenem Ausmaß einzusetzen (Darling, 2017, S. 13; Złotowski, Proudfoot, Yogeewaran & Bartneck, 2015).

Dies gilt insbesondere vor dem Hintergrund, dass Interaktionen zwischen Menschen und vermenschlichten Robotern eine Vielzahl ethischer Fragestellungen aufwerfen (Breazeal et al., 2016, S. 1944), weshalb einige Forschende das Befördern von Vermenschlichung grundsätzlich kritisieren (Darling, 2017, S. 2). Solche Fragestellungen erhalten inzwischen eine erhöhte Aufmerksamkeit innerhalb aktueller Debatten, sind aber gleichwohl im Vergleich zu technischen Fragestellungen deutlich unterrepräsentiert (Riek & Howard, 2014). Beispielsweise stellt sich in Anlehnung an Kants Verrohungsargument, das dieser im Umgang mit Haustieren vorbrachte (Kant, 1924/1991), die Frage, inwieweit ein gewalttätiger Umgang gegenüber belebt wahrgenommenen Robotern unterbunden werden muss, um eine Verrohung der Sitten zu vermeiden (Darling, 2017, S. 8). Die Vermenschlichung kann ferner eine „Entgrenzung“ zwischen Mensch und Roboter befördern (Onnasch et al., 2019, S. 21f.), die sich zum Großteil nicht durch technologische Fortentwicklungen rechtfertigen lässt, sondern sich aus utopischen Zukunftsvisionen in fiktiven Darstellungen und populärwissenschaftlichen sowie technikedeterministischen Technologiediskursen speist (Kehl & Coenen, 2016). Außerdem implizieren komplexe vermenschlichende Zuschreibungen Erwartungen hinsichtlich der künftigen Aktionen der vermenschlichten Entität, die als Erwartungs-Erwartungen wiederum (vermeintliche) Ansprüche gegenüber der vermenschlichenden Person erzeugen. Wenn diese bspw. einen Roboter vermenschlicht und dadurch als soziales Wesen interpretiert, kann sich daran die Erwartung anschließen, dass dieses Wesen wiederum Erwartungen an die vermenschlichende Person stellt. Solche Erwartungs-Erwartungen charakterisieren und strukturieren „im soziologischen Sinne eine soziale Situation“ (Marquardt, 2017, S. 27) und können handlungsleitend wirken, insofern eine Person geneigt sein könnte, ihr Verhalten den mutmaßlichen sozialen Erwartungen des Roboters anzupassen (Waytz, Cacioppo & Epley, 2010), selbst wenn es letztlich bei einer „vereinseitigte[n] soziale[n] Interaktion“ (Marquardt, 2017, S. 28) bleibt. Um die vielfältigen kontextspezifischen Determinanten, Auswirkungen sowie deren Beurteilungsmechanismen erfassen zu können, sind insbesondere interdisziplinäre Forschungsansätze notwendig, die technische, sozialwissenschaftliche, psychologische und ethische Perspektiven integrieren (AI Now, 2016, S. 5; Campa, 2016; Campolo et al., 2017, S. 2).

3.2.4.4 Effekt des unheimlichen Tals

Wie im vorigen Kapitel angedeutet, können die Bewertungen der Auswirkungen von Vermenschlichung ab einem gewissen Grad der Vermenschlichung umschlagen. Dieser Wendepunkt in der Wahrnehmung von vermenschlichten Robotern erinnert an die bekannte und vielfach rezipierte *Theorie des unheimlichen Tals* (*uncanny valley*) nach Mori (1970). Diese postuliert einen talförmigen Verlauf der wahrgenommenen Unheimlichkeit (*eeriness*) eines Roboters in Abhängigkeit von dessen Menschenähnlichkeit. Demnach entstehen in einem bestimmten Bereich der Menschenähnlichkeit besonders starke Gefühle der Unheimlichkeit. Über die Existenz und die Ursachen für diesen Effekt wird in der Forschungsgemeinschaft aktiv und kontrovers diskutiert (Yogeeswaran et al., 2016). Jüngst

wurde der tal förmige Verlauf der Funktion experimentell anhand einer umfassenden Roboterdatenbank bestätigt (B. Kim, Bruce, Brown, de Visser & Phillips, 2020), wie Abbildung 3.2 zeigt. Interessanterweise zeigten sich ähnliche Kurvenverläufe einschließlich eines stark ausgeprägten Minimums auch in Bezug auf das Vertrauen in einen Roboter sowie auf die empfundene Zuneigung zu demselben und die Zuschreibung interner Zustände, wobei diese Zusammenhänge jeweils durch kognitive Einflussfaktoren moderiert werden, wie bspw. die Reflexion über bestehende mentale Repräsentationen (Abubshait, Momen & Wiese, 2017). Neuere Studien legen auch in diesem Kontext die Vermutung nahe, dass die Unheimlichkeitsgefühle besonders stark ausfallen, wenn der Roboter als Konsequenz des Vermenschlichungsprozesses als empfindungsfähig repräsentiert wird. Bei einer bloßen Zuschreibung der Akteur:inschaft waren diese Effekte weniger stark ausgeprägt (Appel, Izydorczyk, Weber, Mara & Lischetzke, 2020).

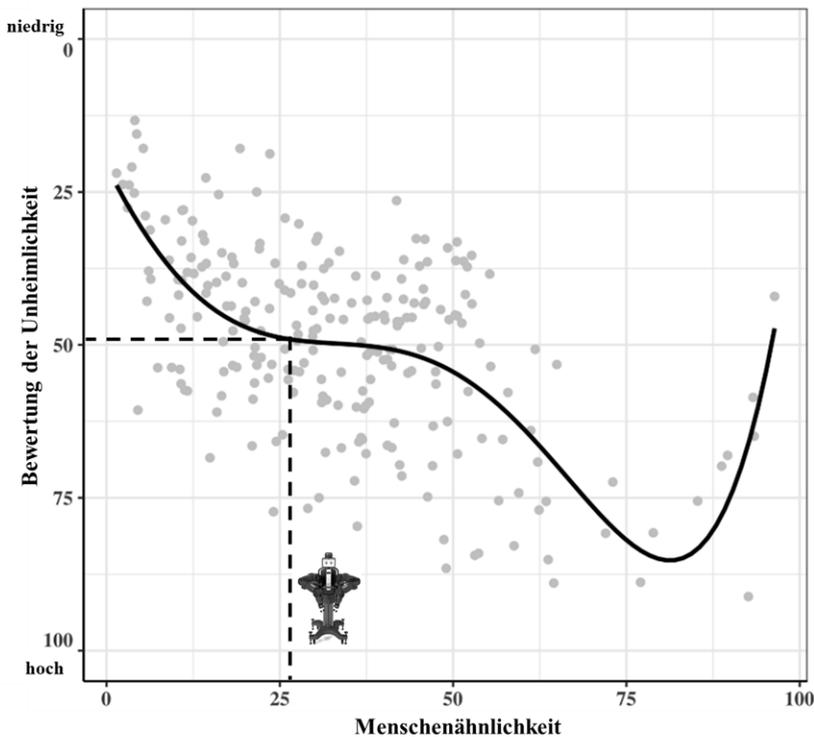


Abbildung 3.2: Unheimlichkeitsbewertungen in Abhängigkeit von der Menschenähnlichkeit von Robotern auf Basis der *Anthropomorphic Robot Database*. Quelle: B. Kim et al. (2020, S. 5), eigene Übersetzung und Einordnung von *Baxter*.

Im industriellen Kontext erscheint die Theorie des unheimlichen Tals derzeit noch weniger relevant zu sein, um die abweichenden Auswirkungen von unterschiedlich starker Vermenschlichung zu erklären, als die zuvor beschriebene Passungshypothese (Goetz et al., 2012), die sich auf die Übereinstimmung von Erwartungen und Beobachtungen bezieht. Denn vermutlich werden industrielle Roboter in den nächsten Jahren keine derart menschenähnliche Gestalt annehmen, dass die Gefahr besteht, starke Unheimlichkeitsgefühle hervorzurufen. So liegt der aktuelle Menschenähnlichkeitswert bspw. für den hybriden Roboter *Baxter* nur bei 27.3 (*The ABOT Database*, 2019), also am Anfang eines Plateaus im Kurvenverlauf (vgl. Abbildung 3.2). Somit ließe sich die Menschenähnlichkeit bis auf einen Score von etwa 45 steigern, ohne dass dieser Cobot unheimlicher wahrgenommen würde. Einschränkend ist hierbei zu erwähnen, dass die verwendeten Menschenähnlichkeitswerte nur auf einer Bewertung des Roboterdesigns basieren und damit von der individuell *wahrgenommenen* Menschenähnlichkeit abweichen können. In diesem Zusammenhang besteht eine Erkenntnislücke in Hinblick darauf, inwiefern menschenähnliche Repräsentationen, die u. a. durch eine vermenschlichte Beschreibung eines Roboters (vgl. Kapitel 3.3) evoziert werden, ebenfalls ab einem gewissen Niveau ein Gefühl der Unheimlichkeit vermitteln.

3.3 Sprachliches Framing

„Indeed, language is an important condition of possibility of our use of technologies”.
 – Coeckelbergh, 2018, S. 150 –

3.3.1 Zur Rolle von Sprache im MRI-Kontext

Wie das einleitende Zitat bereits andeutet, spielt Sprache eine wichtige Rolle bei der Beschreibung, Interpretation und Konstitution von Mensch-Roboter-Relationen (Coeckelbergh, 2011). Einerseits repräsentiert und interpretiert sie Roboter und konstruiert diese andererseits als Quasi-Gegenüber (*quasi-other*) und die Mensch-Roboter-Beziehung als soziale Beziehung (Coeckelbergh, 2010b, 2011, S. 61). Roboter werden innerhalb des sozialen Geschehens zwischen Menschen, also durch die zwischenmenschliche Kommunikation, konstituiert und können daher als ein Gegenstand sozialen Geschehens und ihre Repräsentationen als sozial konstruiert betrachtet werden (Vincent et al., 2015, S. 38). Auf diesen gesellschaftlich verhandelten Konstitutionen basiert die Einstellung gegenüber Robotern, die wiederum für deren Wahrnehmung und Akzeptanz maßgeblich ist. Dieses komplexe Verhältnis zwischen Sprache und Technologie wurde bisher noch unzureichend empirisch erforscht (Coeckelbergh, 2017, S. 189).

Im Folgenden wird zwischen drei unterschiedlichen Szenarien differenziert, in denen Menschen, Roboter und Sprache unterschiedlich zueinander konstelliert sind. Diese Konstellationen lassen sich anhand der jeweiligen Rolle des beteiligten Roboters unterscheiden. Dieser kann erstens ein kommunizierendes Subjekt darstellen, sofern eine Sprachausgabe implementiert ist, zweitens als Adressat:in von Äußerungen eines Menschen auftreten, der mit einem Roboter zu kommunizieren versucht, oder drittens ein Gegenstand der Kommunikation sein, die zwischen Menschen stattfindet (vgl. Abbildung 3.3). Dabei fördert jede Perspektive einen ergänzenden Blick zur Rolle der Sprache in Mensch-Roboter-Interaktionen zutage, der nachfolgend jeweils erläutert wird.

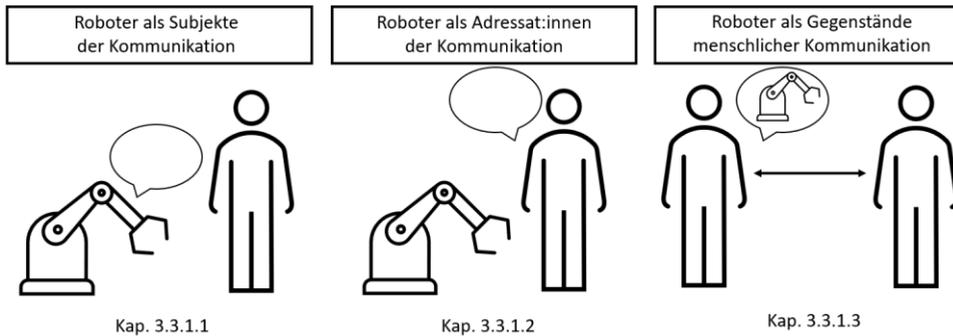


Abbildung 3.3: Verschiedene mögliche Konstellationen von Menschen, Robotern und Sprache.

3.3.1.1 Roboter (mit Sprachausgabe) als Subjekte der Kommunikation

Roboter, die mit einer Funktion zur Ausgabe natürlicher Sprache versehen sind, können als Subjekte der Kommunikation fungieren. Zwar ist dieser Aspekt im Rahmen der vorliegenden Dissertation aus praktischer Sicht derzeit von untergeordneter Relevanz, da Cobots im betrieblichen Umfeld bisher nicht mit einer Sprachausgabe ausgestattet sind, dennoch verdeutlicht er eindrucksvoll die Wirkkraft von Sprache. So begünstigt die Sprachausgabe eines Roboters dessen Vermenschlichung (Schröder & Epley, 2016), insbesondere wenn Roboter auf sich selbst in der ersten Person Singular (*ich*) referieren (Nass & Brave, 2007, S. 119). Die Sprachverarbeitung aktiviert bei Menschen Gehirnareale, die mit sozialem Verhalten assoziiert sind, und provoziert damit unbewusst soziale Reaktionen, die nicht willentlich unterdrückt werden können, wie z. B. Zuschreibungen von Eigenschaften an Sprechende aufgrund von deren Stimme (Nass & Brave, 2007). Dabei ist es unerheblich, ob das Gesprochene von einem Menschen oder einer Maschine produziert wird. Daher sind Sprachschnittstellen „intrinsisch soziale Schnittstellen“ (*intrinsically social interfaces*) (Nass & Brave, 2007, S. 4) und Sprache ist „intrinsisch mit dem Sozialen verbunden“ (*intrinsically connected to the social*) (Coeckelbergh, 2011, S. 62).

Die starke Wirkung sprechender Roboter auf die Wahrnehmung beteiligter Personen demonstrieren auch Studien aus der Entwicklungspsychologie. Normalerweise lernen Kinder neue Phoneme dadurch, dass sie den Eltern oder anderen anwesenden Personen beim Sprechen zuhören. Dieser Lerneffekt bleibt allerdings aus, wenn keine soziale Präsenz der Sprechenden wahrgenommen wird, Kinder also z. B. gesprochene Sprache aus dem Fernseher hören. Die Interaktion mit einem Roboter genügt allerdings, um den Lerneffekt hervorzurufen. Demnach scheint dieser aus Sicht der Kinder die notwendige soziale Präsenz für den Lernprozess aufzuweisen (Kory Westlund et al., 2017, S. 2).

Ferner ist aus linguistischen Studien bekannt, dass Menschen in einer Gruppe ihre Sprache einander angleichen. Dieser Konvergenzeffekt bezieht sich sowohl auf phonetische und syntaktische als auch auf semantische Aspekte (Brandstetter, Beckner, Sandoval & Bartneck, 2017; Lohse, 2012). Brandstetter et al. (2017) fanden in einer Studie heraus, dass sich dieser Effekt auch bei der Interaktion mit Robotern in Bezug auf das verwendete Vokabular einstellt (*lexical entrainment*) und nach der unmittelbaren Interaktion fortbesteht, wenngleich in signifikant schwächerer Form als bei der Interaktion zwischen Menschen. Eine kritische Masse in gleicher Art und Weise kommunizierender Roboter könnte demnach in der Lage sein, menschliche Sprechgewohnheiten zu beeinflussen (Brandstetter & Bartneck, 2017).

3.3.1.2 Roboter als Adressat:innen menschlicher Äußerungen

Menschen neigen dazu, mit technischen Geräten wie bspw. Robotern zu sprechen, unabhängig davon, ob diese mit einer Spracherkennung ausgestattet sind oder nicht (Richert et al., 2017). Mit der zunehmenden Verbreitung sog. Sprachassistenten wird die Kommunikation mit technischen Geräten üblicher. Die Art und Weise, wie eine Person mit einem Roboter spricht, bietet eine aufschlussreiche Informationsbasis, um Einsichten in diejenigen mentalen Modelle der Person zu erlangen, mit denen diese den Roboter und ihre Beziehung zu demselben repräsentiert (Fischer, 2011; Fischer et al., 2012). Je nach Wahrnehmung der aktuellen Interaktionssituation ändern sich die Merkmale der verwendeten Sprache (Lohse, 2009, S. 32, 2012). Hierbei manifestiert sich die Konstitution des Roboters als Quasi-Gegenüber sprachlich durch dessen direkte Ansprache in der zweiten Person (*du*) im Gegensatz zum Sprechen über ihn in der dritten Person (*er/sie/es*)⁴ (Coeckelbergh, 2011, S. 63). Auch weitere interaktionale Sprachmerkmale gelten in diesem Zusammenhang als aussagekräftig, bspw. die direkte Ansprache mit dem Namen oder über Imperative, das Stellen von Verständnisfragen, die Benutzung des inklusiven Pronomens *wir* oder

⁴ Im englischen Sprachraum gilt ferner die Referenz auf einen Roboter mittels der Personalpronomen *he/she* im Vergleich zum neutralen Pronomen *it* als Indiz für die Wahrnehmung des Roboters als Quasi-Gegenüber (s. hierzu Coeckelbergh, 2011).

die Verwendung von Modalpartikeln. Experimentell ließ sich zeigen, dass diese Sprachmerkmale stärker gegenüber verkörperten Robotern auftreten (Fischer et al., 2012).

3.3.1.3 Roboter als Gegenstände menschlicher Kommunikation

Roboter werden zum Gegenstand menschlicher Sprache, wenn Menschen miteinander *über* einen Roboter sprechen. Der Roboter selbst muss hierzu weder Sprache generieren noch verstehen können. Die Art und Weise, wie über Roboter kommuniziert wird, stellt keine reine Repräsentation einer objektiven Realität dar, sondern interpretiert die Relationen zu technischen Artefakten, die als Quasi-Gegenüber erscheinen können (Coeckelbergh, 2011). Ferner lässt die menschliche Rede über Roboter aussagekräftige Rückschlüsse auf vorherrschende Weltanschauungen und Ontologien zu. So ist die Sprache in der westlichen Kultur stark durch einen Subjekt-Objekt-Dualismus geprägt, der wenig Raum für Hybridformen und relationale Ontologien lässt (Coeckelbergh, 2011). In Bezug auf Mensch-Roboter-Relationen hat dies zur Folge, dass Roboter typischerweise als Objekte und Menschen als Subjekte aufgefasst werden. Interaktive Technologien und die in Bezug auf diese verwendete vermenschlichende Sprache fordern den beschriebenen Dualismus heraus, indem sie die Grenzen zwischen Subjekt und Objekt scheinbar verschwimmen lassen (Grunwald, 2019b). Der Umstand, dass es keine Sprache gibt, um über Interaktionspartner:innen zu sprechen, die mehr als nur Dinge sind (*more-than-things*), führen zu einer wahrgenommenen Diskrepanz zwischen dem, was in den Gedanken und mit den Mitteln der gegenwärtigen Sprache fassbar ist auf der einen Seite und unseren affektiven Wahrnehmungen und Gefühlen auf der anderen Seite. Damit einher geht eine Spannung zwischen der sozial akzeptierten Auffassung von Robotern als Maschinen und dazu abweichenden emotionalen Reaktionen, die gegenwärtig weder sozial akzeptiert noch sprachlich verankert sind (Coeckelbergh, 2018, S. 154).

Je mehr Roboter in der Alltagswelt von Menschen als Interaktionspartner:innen auftreten, desto höher ist die Wahrscheinlichkeit, dass sich gewisse sprachliche Metaphern manifestieren werden, um die Aktionen dieser Maschinen zu interpretieren (Richards & Smart, 2016, S. 18ff.). Gängige Metaphern und Narrative prägen wiederum die mentalen Modelle von Robotern und damit die gesellschaftliche Perspektive auf Technologie im Allgemeinen und auf Roboter im Speziellen (Canellas et al., 2017, S. 15). Angesichts dieses vermuteten Einflusses der Sprache fordert Coeckelbergh (2017, S. 189) eine stärkere empirische Untersuchung des Zusammenhangs zwischen Sprache, Technologie und deren Wahrnehmung aus einer phänomenologischen Perspektive. Am Beispiel des Computers verdeutlicht er die Koevolution von Technologien, deren Benutzung und unserer Rede darüber. Während gemäß der etymologischen Bedeutung ein Computer eine Rechenmaschine ist und einst auch so benutzt wurde, wandelte sich das Verständnis hin zu einem persönlichen Gerät (*personal computer*). Heute erscheinen die Begriffe nicht mehr angemessen, da sich die Vorstellung von dieser Technologie sowie die Nutzungsarten und Relationen gewandelt haben.

Computer sind nun keine großen sichtbaren Geräte mehr, sondern sind sowohl augenscheinlich als auch begrifflich verschwunden, bspw. in Gestalt persönlicher Assistenten, Knotenpunkte in einem Netzwerk oder intelligenter Mobilgeräte. Die Rolle der menschlichen Technikverwender:innen wandelt sich ebenfalls sprachlich, insofern gerade in Bezug auf Roboter Nutzer:innen und Operator:innen sprachlich zunehmend zu Interaktionspartner:innen und Personen zu informationsverarbeitenden Einheiten innerhalb technischer Netzwerke werden. Sprache bedingt damit auch stets den Möglichkeitsraum der Technologieverwendung und der Verortung des Menschen im Verhältnis zur Technologie (Coeckelbergh, 2017, S. 187). Desto mehr bedarf es einer genauen Betrachtung der Rede über Roboter und deren Konsequenzen. In diesem Zusammenhang sind Prozesse des sprachlichen Framings von besonderer Relevanz, deren theoretische Ansätze im nächsten Kapitel beschrieben werden.

3.3.2 Theoretische Ansätze

Nachdem im vorigen Kapitel die vielfältige Wirkung von Sprache in verschiedenen Konstellationen zwischen Menschen und Robotern sowie die Wechselwirkungen mit mentalen Modellen verdeutlicht wurden, schließen sich in diesem Kapitel theoretische Ansätze zum sprachlichen Framing an. Zunächst lässt sich festhalten, dass Framing-Prozesse in erster Linie relevant sind, wenn Menschen miteinander über Roboter sprechen, wie in Kapitel 3.3.1.3 beschrieben. Der Begriff des Framing erlangte im Rahmen der frühen Forschungen zur Entscheidungstheorie Popularität als Erklärungsansatz dafür, dass Menschen entgegen voriger Annahmen ihre Entscheidungen nicht vollständig rational treffen (Kahneman & Tversky, 1984; Tversky & Kahneman, 1981, 1986). In diesen Studien zeigte sich, dass für die Abschätzung der Entscheidungskonsequenzen irrelevante Veränderungen an der jeweiligen Ausformulierung des Entscheidungsproblems zu unterschiedlichen Entscheidungen führten, bspw. in Abhängigkeit davon, ob mit der Entscheidung Zugewinne oder Verluste verbunden wurden. Bereits frühzeitig folgerten Tversky und Kahneman (1986, S. 5272), dass die Allgegenwart von Framing-Effekten es unmöglich mache, diese in einer Entscheidungstheorie außen vor zu lassen, die den Anspruch erhebe, eine akkurate Beschreibung menschlicher Entscheidungen zu liefern. In der Folgezeit wurde das Konzept des Framings im Rahmen unterschiedlichster Disziplinen wie der Soziologie, Ökonomie, Psychologie, der kognitiven Linguistik und nicht zuletzt der Kommunikations-, Politik- und Medienwissenschaft angewendet (Borah, 2011, S. 246). Dennoch oder gerade deswegen blieb Framing allerdings ein theoretisch nicht einheitlich gefasstes und „fragmenthaftes Paradigma“ (*fractured paradigm*) (Borah, 2011; Chong & Druckman, 2007; Entman, 1993, S. 51).

Wenngleich der Framing-Begriff theoretisch nur vage bestimmt ist, steht bei den meisten Ansätzen die „automatische Aktivierung kategorial organisierter Wissens- und

Bewertungsstrukturen im Mittelpunkt“ (Stocké, 2002, S. 31). Entman (1993, S. 51ff.) verweist ferner darauf, dass Framing bedeutet, die Salienz selektiver Aspekte einer wahrgenommenen Realität in einem gegebenen Kommunikationskontext zu steigern. Damit legt Framing den Fokus darauf, wie die konkrete Informationsvermittlung die Wahrnehmung einer Situation oder eines Gegenübers beeinflusst. Die in einem Text beinhalteten Frames manifestieren sich „durch die Präsenz oder Absenz bestimmter Schlüsselwörter, Standardphrasen, stereotyper Bilder, Informationsquellen und Sätze“ (Entman, 1993, S. 52; Übersetzung des Verfassers), die bestehende Denkmuster ansprechen und hervorheben. Framing-Prozesse können dazu beitragen, dass bestimmte mentale Modelle situativ für die Bewältigung einer gewissen Situation herangezogen werden. Zwar existiert explizit eine Theorie der mentalen Modelle (*mental model theory*) zur Erklärung von Framing-Effekten (Stocké, 2002, S. 62), allerdings verweisen verschiedene theoretische Ansätze zum Framing auf unterschiedliche kognitive Konstrukte wie Einstellungen, Stereotype, Skripte, Rollen oder Schemata (Entman, 1993, S. 53; Stocké, 2002, S. 76). Dabei spiegelt sich erneut die in Kapitel 3.1.3 angedeutete definitorische Unschärfe in Bezug auf mentale Modelle und verwandte mentale Repräsentationsformen wider.

Grundsätzlich lässt sich von Framing sprechen⁵, wenn die Formulierungen der Sprechenden mentale Modelle bei der rezipierenden Person als Ergebnis eines Bottom-Up-Verarbeitungsprozesses aktivieren und/oder modifizieren. Die hervorgerufenen Veränderungen auf hoher Verarbeitungsebene führen dann wiederum in einem Top-Down-Prozess dazu, dass wahrgenommene Situationen vor dem Hintergrund einer gewissen gedanklichen Struktur interpretiert werden, die sich in den mentalen Modellen widerspiegelt (Groom et al., 2011, S. 195; Kory & Kleinberger, 2014, S. 9; Stenzel, Chinellato, del Pobil et al., 2012; Stenzel, Chinellato, Bou et al., 2012). Der Vorgang des Framings basiert somit primär auf der Bottom-Up-Verarbeitung von in bestimmter Weise gerahmten wahrgenommenen Informationen. Allerdings greifen hierbei bereits Bottom-Up-Reize und Top-Down-Erwartungen ineinander, sodass die Rezeption eines Framing-Textes stets auch eine Interaktion mit bereits bestehenden mentalen Modellen darstellt (Entman, 1993, S. 53; Klapper, Ramsey, Wigboldus & Cross, 2014, S. 2503). Die Anwendung dieser in mentale Modelle integrierten gerahmten Informationen geschieht zu einem späteren Zeitpunkt primär in einem Top-Down-Prozess. Daran verdeutlicht sich der enge Zusammenhang zwischen Framing-Prozessen und Erwartungen, sodass sich Framing auch als eine Form des Erwartungsmanagements interpretieren lässt (Kory & Kleinberger, 2014, S. 9). Dabei kann das Verhältnis aus der kognitiv weniger aufwändigen und modellbasierten Top-Down-Verarbeitung und der aufwändigen datenbasierten Bottom-Up-Verarbeitung je nach Bereitschaft zur kognitiven Anstrengung variieren (Stocké, 2002, S. 84). Unter Umständen können sich Informationsgehalte, die aus mentalen Modellen entstammen und durch ein früheres

⁵ Mitunter wird der Begriff des Priming als mögliches Synonym angesehen, vgl. Chong und Druckman (2007, S. 114f.).

Framing entstanden sind, gegenüber den situativ verfügbaren und wahrgenommenen Hinweisreizen durchsetzen, sodass durch Framing induzierte Erwartungen nicht nur mangelnde Umweltinformationen ergänzen, sondern diese auch überschreiben können (Stocké, 2002, S. 94).

Wie Studien aus Forschungsgebieten außerhalb des MRI-Kontexts zeigen, können bereits vergleichsweise kleine Variationen in der Wortverwendung unterschiedliche Assoziationen und Verhaltensänderungen zur Folge haben (Brandstetter, 2017; Dijksterhuis & Bargh, 2001). Aus der Entscheidungstheorie ist bekannt, „dass oft minimale Veränderungen in der Art der Informationsvermittlung und geringfügige Variationen des Entscheidungskontextes zu oft dramatischen Veränderungen im Entscheidungsverhalten führen“ (Stocké, 2002, S. 10). Dabei bezogen sich die klassischen Forschungen von Kahneman und Tversky (1984) in erster Linie auf sog. Äquivalenz-Frames (*equivalence frames*), bei denen eindeutig äquivalente Informationen wie komplementäre Wahrscheinlichkeitswerte verwendet werden, etwa wenn z. B. bei einer dichotomen Entscheidung entweder darauf verwiesen wird, dass eine 40-prozentige Chance besteht, dass Entscheidung 1 richtig ist, oder dass eine 60-prozentige Chance besteht, dass Entscheidung 2 richtig ist. Konzeptionell ist diese Art von Framing zu unterscheiden von sog. Hervorhebungsframes (*emphasis frames*), bei denen sprachlich bestimmte Aspekte besonders akzentuiert werden, bspw. durch die Verwendung unterschiedlich konnotierter Begrifflichkeiten. In der Praxis sind die Unterscheidungen allerdings mitunter schwer zu bestimmen (Chong & Druckman, 2007, S. 114). Bei der Untersuchung von Informationen in den Medien und in öffentlichen Debatten stehen in der Regel Hervorhebungsframes im Fokus, die wiederum entweder gegenstandsspezifisch (*issue-specific*) oder generisch (*generic*) sein können (Lecheler & de Vreese, 2011, S. 961).

Die durch Framing evozierten Veränderungen können mitunter subtil und kurzlebig sein, sodass sie sich nicht mittels bewusster Selbstauskünfte erheben, sondern nur beobachten lassen (Dijksterhuis & Bargh, 2001; Kory & Kleinberger, 2014; Kory Westlund et al., 2016, S. 689). So liefen bspw. Versuchspersonen langsamer, wenn vorher ihre mentalen Modelle von älteren Personen aktiviert wurden (Bargh, Chen & Burrows, 1996). In populärwissenschaftlichen Publikationen wurde jüngst die gesellschaftliche Dimension solcher Framing-Prozesse thematisiert, die sich über Narrative als Bündel von Frames signifikant auf gesellschaftliche und politische Debatten auswirken können (Leifeld & Haunss, 2012, S. 384)⁶. Hieran zeigt sich, dass Framing sowohl soziologische als auch psychologische Aspekte besitzt. Erstere beziehen sich auf kollektive Deutungsrahmen, die über populäre virulente Frames implizit als Verständigungsmechanismen kommuniziert werden, und Letztere auf den Einfluss der individuellen Umweltwahrnehmung (Borah, 2011, S. 247f.).

⁶ Zur Relevanz des politischen Framings vergleiche z. B. Wehling (2016), zur Relevanz intersubjektiver Narrative Harari (2015) oder zu Framing im Bereich der Markenforschung Newman (2018, S. 514f.).

In Hinblick auf die Technologieakzeptanz ergeben sich aus unternehmerischer Sicht Potenziale, um mittels gezieltem sprachlichen Framing während eines Robotereinführungsprozesses etwaigen Ressentiments unter den betroffenen Mitarbeitenden entgegenzuwirken (Charalambous, Fletcher & Webb, 2013, S. 302).

Stocké (2002) klassifiziert im Rahmen einer umfassenden Analyse die Vielzahl bestehender Framing-Ansätze in drei Typen, die erstens durch Ambiguitätseinflüsse, zweitens durch Heuristikverwendung und drittens durch Schemaaktivierung charakterisiert sind⁷. Bei ambiguitätsorientierten Framing-Ansätzen steht das Fehlen entscheidungsrelevanter Informationen im Vordergrund, die im Rahmen unterschiedlicher Framing-Bedingungen verschiedenartig ergänzt werden. Heuristikbasierte Framing-Ansätze beruhen auf der Annahme, dass Personen etwa aus Gründen mangelnder Motivation oder kognitiver Ökonomie nur besonders saliente Oberflächenmerkmale von Umgebungsvariablen für ihre Entscheidung heranziehen. Diese situative Salienz lässt sich wiederum durch Framing beeinflussen. Schemabasierte Ansätze gehen davon aus, dass in Abhängigkeit bestehender Hinweisreize unterschiedliche Schemata aktiviert werden, was die Wahrnehmung der Situation beeinflusst (Stocké, 2002, S. 35). Dieser Ansatz operiert explizit mit dem Begriff des (kognitiven) Schemas. Unter Rückgriff auf die Erläuterungen zu mentalen Modellen in Kapitel 3.1 und zu möglichen Abgrenzungen sowie Zusammenhängen mit Schemata in Kapitel 3.1.3 ist davon auszugehen, dass sich ein mentales Modell der wahrgenommenen Situation auf Basis der aktivierten Schemata bildet. Die Festlegungen innerhalb dieses Modells überwinden die „Mehrdeutigkeit und Interpretationsbedürftigkeit objektiver Wahrnehmungsdaten“ (Stocké, 2002, S. 69) und wirken handlungsleitend. Die Ausgestaltung verändert sich dabei durch „geringfügige und vom Standpunkt instrumenteller Rationalität irrelevante Unterschiede [in der Informationspräsentation] zwischen den Framing-Bedingungen“ (Stocké, 2002, S. 68/83). Da die Anwendung mentaler Modelle von einer Entscheidungssituation auch als heuristische Herangehensweise interpretiert werden kann, zeigen sich bei subjektiv als sehr wichtig empfundenen Entscheidungen nur sehr geringe formulierungsbasierte Framing-Effekte. In diesen Fällen wird eine tiefere, nicht rein heuristikbasierte Verarbeitung herbeigeführt (Stocké, 2002, S. 144), was einen Zusammenhang zwischen Framing-Effekten und verschiedenen Verarbeitungsmodi andeutet.

Innerhalb der vorliegenden Dissertation bezeichnet das sprachliche Framing den Vorgang, bei dem faktisch identische Informationen durch die Wahl unterschiedlicher sprachlicher Merkmale wie der Verwendung bestimmter Formulierungen oder Satzstrukturen von der rezipierenden Person innerhalb unterschiedlicher Deutungsrahmen verstanden werden und damit unterschiedliche mentale Modelle aktivieren. Stocké (2002, S. 23f.) spricht in diesem Zusammenhang von „Framing-Effekte[n] im strikten Sinne“, die auf „Unterschiede in

⁷ Für eine tabellarische Übersicht siehe Stocké (2002, S. 93).

der Oberflächenstruktur“ zurückgehen, welche sich wiederum nur aus der unterschiedlichen Darstellung des exakt gleichen Entscheidungsproblems ergeben.

3.3.3 Sprachliches Framing von Robotern

Die Roboterforschung fokussierte sich in der Vergangenheit beinahe ausschließlich auf den Einfluss von wahrnehmbaren Robotereigenschaften auf das Vertrauen, sodass viele Studien sichtbare Merkmale wie das anthropomorphe Design manipulierten und die Auswirkungen dieser Veränderung im Rahmen von Bottom-Up-Verarbeitungsprozessen beobachteten (Coeckelbergh, 2011; Groom et al., 2011). Coeckelbergh (2011, S. 65) regte bereits vor einem Jahrzehnt an, den üblichen Forschungsansatz der Untersuchung von Bottom-Up-Prozessen umzudrehen und stattdessen die Auswirkungen eines experimentell veränderten sprachlichen Umfelds (*linguistic environment*) zu untersuchen. In den vergangenen Jahren entstanden erste Untersuchungen zum Framing von Robotern, die allerdings gegenüber den Studien zu Bottom-Up-Einflüssen weiterhin stark unterrepräsentiert sind.

Ein Großteil dieser anwendungsnahen MRI-Studien beziehen sich zwar auf den Framing-Begriff, legen dabei allerdings mitunter ein sehr weit gefasstes Verständnis von Framing zugrunde, das signifikant über die zuvor beschriebene Definition des strikten Framings hinausgeht. Bei derartigen Studien werden den Teilnehmenden unterschiedliche Informationen über einen Roboter vermittelt und dadurch verschiedene Erwartungen evoziert. Die experimentelle Manipulation erfolgt also nicht subtil über die Variation sprachlicher Merkmale, sondern wesentlich expliziter über die Veränderung der Faktenlage. So manipulierten bspw. Rueben, Bernieri, Grimm & Smart (2017) in einer Untersuchung zu Datenschutzbedenken in Bezug auf einen Haushaltsroboter die Angabe darüber, ob dieser im eigenen häuslichen Umfeld der Versuchspersonen entweder von einer vertrauten Person wie der eigenen Schwester oder einer unbekannt Person wie einem Hausgutachter ferngesteuert wird. Paepcke & Takayama (2010) variierten in einem Experiment mit den Robotern *Pleo* und *Aibo* die Angabe, ob der jeweilige Roboter über interaktive Fähigkeiten zur Erkennung von Menschen verfügt oder nicht. Ähnlich verhält es sich in einer Studie zu Framing-Effekten bei Rettungsrobotern von Groom et al. (2011). Dabei wurde den Versuchsteilnehmenden im Vorfeld entweder mitgeteilt, dass der Roboter nur ein Medium darstellt, um mit einem menschlichen Rettungsteam kommunizieren zu können, oder dass der Roboter selbstständig Hilfestellung leistet (Groom et al., 2011). Entsprechend unterschiedliche Erwartungen der Versuchsteilnehmenden an die Fähigkeiten der verwendeten Roboter lassen sich in diesen Studien eher als offensichtliche Folge der unterschiedlichen Faktenlage denn als Indiz für die Wirkkraft sprachlich vermittelter Frames interpretieren.

Gleichwohl lassen sich einige Studien identifizieren, die die Effekte von verschiedenen Arten des Framing auf die Wahrnehmung eines Roboters bzw. einer MRI empirisch

untersuchen. Dabei wird häufig die wahrgenommene Menschenähnlichkeit des Roboters durch Framing manipuliert. So untersuchten Nijssen, Müller, van Baaren und Paulus (2019) in einer Serie von zwei quantitativen Experimenten, inwiefern das menschenähnliche Framing eines Roboters dessen Berücksichtigung in moralischen Dilemma-Situationen verändert. Dabei zeigte sich, dass ein menschenähnlich geframter Roboter mit einer geringeren Wahrscheinlichkeit im Rahmen des moralischen Dilemmas geopfert wird. Offenbar schreiben Menschen dem Roboter in Abhängigkeit vom jeweiligen Framing affektive Zustände zu und reagieren empathischer auf ihn. Wenn Menschen diesen empathischen Blickwinkel einnehmen, neigen sie eher dazu, andere Personen zu verschonen (Majdandžić et al., 2012). In einem viel beachteten Experiment fanden Darling, Nandy & Breazeal (2015) ferner heraus, dass Versuchspersonen länger zögerten, einen insektenähnlichen Roboter (*hexbug*) mit einem Hammer zu zerstören, wenn sie zuvor eine Hintergrundgeschichte zu diesem Roboter gehört hatten und aufgrund dessen mehr Empathie gegenüber dem Roboter empfanden. Überdies scheint die Beschreibung eines Roboters als biologisches, menschenähnliches Wesen dazu zu führen, dass Menschen die Aktionen des Roboters unterbewusst mental ko-repräsentieren (Stenzel, Chinellato, Bou et al., 2012). Dazu passt die Erkenntnis, dass sich anthropomorphe Beschreibungen eines Roboters positiv auf das Erinnerungsvermögen an Ereignisse aus Narrativen über den Roboter auswirken (L. J. Baker et al., 2018). Während die bisher beschriebenen Befunde auf eine intensivere Verarbeitung von Informationen über vermenschlichte Entitäten und ein höheres Maß an Einfühlung hinweisen, zeigt eine Studie von Wen et al. (2018), dass Personen die Hilfestellungen eines maschinell beschriebenen Roboters signifikant häufiger akzeptierten und weniger lange zögerten, bis sie diese Entscheidung fällten.

Des Weiteren ist die Effektivität von sprachlichem Framing bei Kindern gut belegt. In einer entsprechenden Studie zeigte sich, dass Kinder signifikant mehr mit einem Kuschelroboter sprachen und sozial interagierten, wenn dieser zuvor von der Versuchsleitung persönlich angesprochen und menschenähnlich beschrieben wurde. Da sich keine signifikanten framingabhängigen Unterschiede in Bezug auf abgefragte Eigenschaften des Roboters finden ließen, erachten die Autor:innen das beobachtbare Sozialverhalten in Übereinstimmung mit früheren Studien als unbewussten Prozess (Kory & Kleinberger, 2014). Auch in weiteren Studien mit Kindern zeigten sich subtile Unterschiede im Blickverhalten, der Schüchternheit sowie im Erinnerungsvermögen, je nachdem, ob der Roboter als soziale:r Akteur:in oder maschinenähnlich geframt wurde (Kory Westlund et al., 2016; Kory Westlund et al., 2017). Außerdem wirkte sich in einer weiteren Studie u. a. das Hinzufügen einer Hintergrundgeschichte auf den empfundenen Spaß während der Interaktion, die Wahrnehmung der Beziehung zum Roboter und die Fähigkeit, eine Geschichte nachzuerzählen, aus (Kory Westlund & Breazeal, 2019).

Insgesamt ergibt sich in der Gesamtschau eine Summe experimenteller Hinweise auf die Wirksamkeit von Framing auf verschiedene Dimensionen der Wahrnehmung eines

Roboters. Teilweise führten bereits minimale Variationen in den sprachlichen Merkmalen zu nachweisbaren Effekten. Diese scheinen sich eher in verhaltensorientierten Maßen widerzuspiegeln als in Selbstauskünften (Stenzel, Chinellato, Bou et al., 2012), was sich beziehend auf die Theorie des dualen Anthropomorphismus (vgl. Kapitel 3.2.3.2) als Ausdruck einer impliziten Vermenschlichung verstehen lässt. Framing scheint demnach unbewusst zu wirken. Alle Studien haben gemein, dass das Framing rein textbasiert vorgenommen wurde, ohne dass eine reale MRI stattfand. Außerdem ist die Anwendbarkeit auf den MRI-Kontext größtenteils stark eingeschränkt, da die Stichproben aus Kindern oder Studierenden bestanden oder höchst fiktive Anwendungsfälle wie das willentliche Zerstören insektenähnlicher Roboter betrachtet wurden (Onnasch & Roesler, 2019).

Aufgrund dieser methodischen Schwächen verfolgten Onnasch und Roesler (2019) das Ziel, Framing-Effekte in einem realitätsnäheren Szenario zu untersuchen. In einer entsprechenden Studie erhielten die teilnehmenden Studierenden textuelle Beschreibungen eines Roboters, die entweder menschenähnlich (*er*, Benennung einer Lieblingsfarbe, etc.) oder maschinell mit technischen Spezifikationen gestaltet waren. Daraufhin sollten die Studierenden in einem kooperativen Interaktionsszenario mit dem humanoiden Roboter *Nao* den sog. *Turm von Hanoi* aufbauen, indem sie dessen Handlungsanweisungen folgten. Es zeigte sich allerdings, dass die Versuchsteilnehmenden den anthropomorph geframten Roboter nicht als signifikant menschenähnlicher wahrnahmen und sogar in geringerem Umfang dazu bereit waren, für eine vorgegebene anstehende Reparatur des Roboters zu spenden. Die Autorinnen vermuteten, dass der funktionelle Wert des Roboters für die Aufgabe durch das menschenähnliche Framing in den Hintergrund rückt. In einer Replikation des Experiments fügten sie den Framing-Texten deshalb einen Abschnitt hinzu, der diesen Wert unterstrich. Allerdings ergab sich erneut weder eine signifikant höhere wahrgenommene Menschenähnlichkeit noch eine höhere Spendenbereitschaft. Die Autorinnen schließen aus diesen überraschenden Ergebnissen, dass sich positive Framing-Effekte aus der Forschung mit nicht-humanoiden Robotern nicht auf humanoide Roboter wie *Nao* übertragen lassen, da diese allein durch ihr Erscheinungsbild bereits menschenähnlich wirken. Außerdem sei der funktionale Wert von Robotern für die zu lösende Aufgabe essenziell (Onnasch & Roesler, 2019). Insbesondere die ungewöhnliche hierarchische Konstellation mit einem weisungsbefugten Roboter und der Rückgriff auf Studierende als Stichprobe schränken die Validität dieser Studie allerdings ebenfalls ein, sodass die empirische Evidenz für das Vorliegen von Framing-Effekten im Rahmen realistischer und praxisnaher MRI-Anwendungen insgesamt als sehr spärlich zu bezeichnen ist und auf weiteren Forschungsbedarf hinweist. Dies gilt insbesondere auch für die Dauer der Wirkung von Framing-Effekten, die noch nicht hinreichend untersucht ist (Kory & Kleinberger, 2014; Kory Westlund et al., 2016). Während das Forschungsinteresse der bisherigen Studien auf den Auswirkungen von Framing auf ausgewählte Einzelmerkmale der menschlichen Wahrnehmung eines Roboters lag, fokussiert die vorliegende Dissertation den darüber

hinausgehenden mittelbaren Einfluss von Framing auf das relevante Konstrukt des Mensch-Roboter-Vertrauens, das im folgenden Kapitel näher ausgearbeitet wird.

3.4 Mensch-Roboter-Vertrauen

„Der Mensch hat zwar in vielen Situationen die Wahl, ob er in bestimmten Hinsichten Vertrauen schenken will oder nicht. Ohne jegliches Vertrauen aber könnte er morgens sein Bett nicht verlassen. Unbestimmte Angst, lähmendes Entsetzen befiehlt ihn. (...) Solch eine unvermittelte Konfrontierung mit der äußersten Komplexität der Welt hält kein Mensch aus.“

– Luhmann (1968/2014), S. 1 –

3.4.1 Allgemeine Merkmale und Funktionen

Dieses Zitat verdeutlicht den immensen Stellenwert, den Vertrauen im menschlichen Zusammenleben einnimmt. Das Konzept des Vertrauens wird in zahlreichen Kontexten je unterschiedlich definiert, diskutiert und konzeptualisiert, ohne dass ein einheitlicher Konsens für eine Definition besteht (Hoff & Bashir, 2015, S. 409). In diesem Unterkapitel soll in einem ersten Schritt das Konzept des Vertrauens im Allgemeinen begrifflich gefasst und wesentliche Merkmale herausgearbeitet werden (Kapitel 3.4.1). Hierbei fließen neben psychologischen auch soziologische Überlegungen ein, insbesondere aus Niklas Luhmanns Monografie *Vertrauen* (Luhmann, 1968/2014), die sich diesem Themenkomplex ausführlich widmet und auf die sich auch neuere Forschungsarbeiten beziehen (u. a. B. French, Duenser & Heathcote, 2018; Lewicki, Tomlinson & Gillespie, 2006; Sumpff, 2019). Daraufhin folgt im nächsten Unterkapitel eine Betrachtung von Vertrauen als mentaler Zustand (Kapitel 3.4.2) sowie eine Gegenüberstellung von bestehenden Vertrauensdefinitionen im Bereich der Mensch-Technik-Interaktion und eine Weiterentwicklung derselben in Bezug auf die MRI (Kapitel 3.4.3). Daran schließt sich eine Gegenüberstellung von kognitiven und affektiven Vertrauenskomponenten an (Kapitel 3.4.4) sowie eine Betrachtung möglicher vertrauensnehmender Instanzen im Kontext des Mensch-Technik-Vertrauens (Kapitel 3.4.5). Das darauffolgende Kapitel beschäftigt sich mit den Auswirkungen verschiedener Vertrauensniveaus auf die Technologienutzung im Allgemeinen und auf MRI-Situationen im Speziellen (Kapitel 3.4.6). Anschließend erfolgt in Kapitel 3.4.7 die Darstellung dreier einflussreicher theoretischer Modelle zum Vertrauen in Automationstechnologie, die in einem integrierten Vertrauensmodell zusammengeführt werden (Kapitel 3.4.7.4), anhand dessen sich der Fokus dieser Dissertation stärker eingrenzen lässt. Kapitel 3.4.8 beinhaltet Erläuterungen zu den Unterschieden und Gemeinsamkeiten zwischen Vertrauen und dem Konstrukt der Technologieakzeptanz bzw. zwischen Vertrauens- und Technologieakzeptanzmodellen. Daran schließt sich ein Kapitel zur Frage der

Messbarkeit und zu gängigen Messmethoden von Vertrauen (Kapitel 3.4.9) an, bevor im letzten Unterkapitel das Konzept des Misstrauens (Kapitel 3.4.10) thematisiert wird. Diese einleitenden Ausführungen verdeutlichen bereits, dass das Konzept des Vertrauens aus verschiedenen Blickwinkeln beleuchtet wird, um den Mangel an speziell für den MRI-Kontext entwickelten Konzepten und Modellen durch eine Übertragung von allgemeinen Ansätzen aus der Psychologie und Soziologie sowie aus der übergreifenden Mensch-Technik- und der angrenzenden Automationsforschung zu kompensieren. Notwendige Adaptionen für die Übertragung der bestehenden Ansätze und die Grenzen der Übertragbarkeit werden dabei jeweils berücksichtigt und verdeutlicht.

Zunächst stehen in diesem Unterkapitel allerdings die allgemeinen Merkmale von Vertrauen und dessen Funktionen im Vordergrund. Vertrauen ist ein „elementarer Tatbestand des sozialen Lebens“ (Luhmann, 1968/2014, S. 1), der in Hinblick auf die Komplexität desselben beinahe unverzichtbar scheint: „Sind Chaos und lähmende Angst die einzige Alternative zum Vertrauen, so läßt sich daraus folgern, der Mensch solle, seinem Wesen entsprechend, Vertrauen schenken, wenn auch nicht blindlings und nicht in jeder Hinsicht“ (Luhmann, 1968/2014, S. 1). In erster Linie handelt es sich bei Vertrauen um ein soziales Phänomen, denn die Notwendigkeit zum (interpersonalen) Vertrauen entfiel, wenn keine anderen Personen zugegen wären. Gleichzeitig stellt der Prozess des Vertrauens auch einen Gegenstand intrapersonaler psychologischer Untersuchungen dar (Sumpf, 2019, S. 42). In diesem Sinne lässt sich Vertrauen gleichzeitig als multidimensionales psychologisches Konstrukt verstehen (Lewis, Sycara & Walker, 2018, S. 149).

Eine allgemeingültige Definition von Vertrauen ist nur schwerlich möglich (Hoff & Bashir, 2015). Luhmann (1968/2014, S. 1) bestimmt Vertrauen allgemein als „Zutrauen[] zu eigenen Erwartungen“ und verweist hierbei bereits darauf, dass die Erforschung von Vertrauen erfordert, sich mit bestehenden Erwartungen gegenüber Adressat:innen des Vertrauens zu beschäftigen (Sumpf, 2019, S. 31f.). Funktional lässt sich Vertrauen als Mechanismus der Komplexitätsreduktion bestimmen. Die Komplexität ergibt sich hierbei aus der Tatsache, dass sich nicht jede mögliche Erwartung erfüllen kann, oder elementarer formuliert, dass „nicht alle Zukunft Gegenwart (...) werden kann“ (Luhmann, 1968/2014, S. 14). Dieses Leben in Angesicht einer „stets überkomplexen Zukunft“ (Luhmann, 1968/2014, S. 14) stellt eine Überforderung des Menschen dar und liefert die zentrale Motivation dafür, mittels Vertrauen Komplexität zu reduzieren. Zu vertrauen bedeutet, eine „riskante[] Vorleistung“ (Luhmann, 1968/2014, S. 53) zu wagen. Dabei lässt sich die Entscheidung zum Vertrauen nicht vollständig sachlich begründen, da sie stets „durch Überziehen der vorhandenen Information zustande“ kommt und auf „eine[r] Mischung aus Wissen und Nichtwissen“ (Luhmann, 1968/2014, S. 31) gründet. Wäre jedes Nichtwissen über die Zukunft beseitigt, so wäre Vertrauen unnötig; gäbe es kein Wissen im Sinne von Anhaltspunkten über die Wahrscheinlichkeiten des Eintretens gewisser Zukunftsereignisse, entbehrte das Vertrauen jeder Grundlage. Die Prozesse der Vertrauensbildung benötigen

gewisse Anhaltspunkte, die aus vergangenen Erfahrungen abgeleitet werden. Mit der Vergangenheit, die stets bereits eine „reduzierte Komplexität“ (Luhmann, 1968/2014, S. 23) in Abwesenheit möglicher Alternativszenarien darstellt, herrscht eine gewisse Vertrautheit vor. Anhaltspunkte für Vertrauen aus der Vergangenheit abzuleiten, basiert auf der irrigen Annahme, dass sich diese vertrauten Umstände schlicht in gleicher Form in die Zukunft übertragen lassen: „Die Orientierung am Gewesenen kann daher die Welt vereinfachen und verharmlosen. Man unterstellt, daß das Vertraute bleiben, das Bewährte sich wiederholen, die bekannte Welt sich in die Zukunft hinein fortsetzen wird. (...) Man kann nicht ohne jeden Anhaltspunkt und ohne alle Vorerfahrungen Vertrauen schenken. Aber Vertrauen ist keine Folgerung aus der Vergangenheit, sondern es überzieht die Informationen, die es aus der Vergangenheit besitzt und riskiert eine Bestimmung der Zukunft“ (Luhmann, 1968/2014, S. 23f.). Diese Herangehensweise reduziert die Komplexität erheblich, wenngleich sie auf einem unzulässigen induktiven Schluss von der Vergangenheit auf die Zukunft basiert.

Damit sind die wesentlichen Aspekte, die Vertrauen begleiten, bereits bestimmt (Sumpf, 2019, S. 69). Vertrauen reduziert die *Komplexität*, der sich Menschen im Alltag aufgrund der Existenz vieler möglicher Zukünfte gegenübersehen. Die Notwendigkeit zu vertrauen, entsteht demnach aus dem *Nichtwissen* über die Zukunft und das Vertrauen selbst basiert auf diesem Nichtwissen. Das Schenken von Vertrauen geht stets mit dem *Risiko* enttäuschter *Erwartungen* einher. Vertrauen ist der Prozess der *Aufhebung* (*suspension*) widersprüchlicher Zukunftsszenarien und der Überwindung damit verbundener Unsicherheiten und Vulnerabilitäten, indem die oder der Vertrauende so handelt, als seien diese schlicht nicht länger vorhanden (Möllering, 2006). Besonders interessant ist ferner das Verhältnis von Vertrauen zu Kontrolle, die durch das Vertrauen entbehrlich wird. Wenn man alles kontrollieren könnte, bräuchte es kein Vertrauen, während es in einem Zustand völligen Vertrauens *vice versa* keiner Kontrolle mehr bedürfte (Sumpf, 2019, S. 70). Schon die sprichwörtliche Feststellung, dass Vertrauen zwar gut, Kontrolle aber besser sei, konzeptioniert diese beiden Verhaltensweisen als Gegensätze. Allerdings steht hierbei zu bedenken, dass sich Vertrauen und Kontrolle nicht komplett ausschließen. Eine vertrauensvolle Interaktion kommt vielmehr u. U. erst durch die Kombination eines (geringen) Maßes an Vertrauen mit einem gewissen Maß an Kontrolle zustande, wenn die oder der Vertrauensgebende gerade angesichts der gegebenen Kontrollmöglichkeit erwartet, nicht durch die vertrauensnehmende Instanz ausgebeutet zu werden (Beckmann et al., 2004). Die steigende Relevanz von Vertrauen (oder alternativ von Kontrollmechanismen) lässt sich mit einer Zunahme an Komplexität und Freiheitsgraden begründen, die im Unternehmensumfeld aus weniger hierarchischen und vorstrukturierenden organisationalen Führungsstilen sowie aus der Interaktion mit modernen Technologien resultiert, die vielfältige, scheinbar nicht-determinierte Verhaltensweisen an den Tag legen (J. D. Lee & See, 2004, S. 52).

Aufgrund der verschiedenartigsten Betrachtungsperspektiven auf Vertrauen und der Vielzahl unterschiedlicher Definitionen lassen sich nur wenige übergreifende Merkmale von Vertrauen identifizieren. Laut B. French et al. (2018, S. 19) können die folgenden drei Merkmale als Gemeinsamkeiten aller Vertrauensbegriffe in den unterschiedlichen Disziplinen angesehen werden:

1. Vertrauen involviert mindestens zwei Parteien⁸, die in Hinblick auf ihre mentalen Zustände und Aktionen nicht vollständig determiniert sind (Möllering, 2006, S. 3). Eine vertrauensgebende Partei (*trustor*) schenkt sein Vertrauen einer vertrauensnehmenden Partei (*trustee*). Wenn es sich bei beiden Parteien um (unterschiedliche) Personen handelt, wird von interpersonalem Vertrauen gesprochen. Wenn eine Person äußert, in eine Technologie zu vertrauen, dann ist alltagssprachlich von Mensch-Technik-Vertrauen die Rede.⁹ Undenkbar ist es gegenwärtig, dass eine Technologie die Rolle der vertrauensgebenden Partei einnimmt, da dies komplexe mentale Zustände und soziale Überlegungen erfordern würde.
2. Vertrauen ist relevant in Situationen, in denen sich die vertrauende Partei mit Unsicherheit und Vulnerabilität konfrontiert sieht.
3. Vertrauen stellt eine Entscheidungsheuristik dar, die zur Anwendung kommt, wenn ein vollständiges Verständnis der Situation, der Entscheidungsoptionen und der möglichen zukünftigen Folgen dieser Entscheidungen praktisch unmöglich ist.

Ob sich eine Person dazu entscheidet, jemandem Vertrauen entgegenzubringen, hängt davon ab, ob sie die vertrauensnehmende Instanz als vertrauenswürdig einschätzt, also erwartet, dass diese das Vertrauen in einer gegebenen Situation der Unsicherheit nicht missbrauchen wird (Lewis et al., 2018, S. 149). Denn wer vertraut, begibt sich stets in die Gefahr der Ausbeutung durch die Person, der vertraut wird, und macht sich insofern von dieser abhängig (Luhmann, 1968/2014, S. 53). Damit lässt sich Vertrauen auch „als situative Nichtausbeutungserwartung des Vertrauensgebers“ (Beckmann et al., 2004, S. 2) in Bezug auf eine geplante riskante Handlung charakterisieren. Die Zuschreibung von Vertrauenswürdigkeit ist an die Annahmen geknüpft, dass ein Gegenüber über gewisse Fähigkeiten verfügt (*ability*), um die erwartete Handlung auszuführen, der vertrauenden Person wohlwollend gegenübersteht (*benevolence*) und integer ist (*integrity*), also sein Verhalten an einer Menge übergeordneter Prinzipien ausrichtet, die die vertrauensgebende Person goutiert oder zumindest akzeptiert (Mayer, Davis & Schoorman, 1995). Abgesehen von diesen Eigenschaften, die der vertrauensnehmenden Instanz zugeschrieben werden, sind für das tatsächliche Zustandekommen von Vertrauen die (wahrgenommene)

⁸ Das Phänomen des Selbstvertrauens mag an dieser Stelle eine Ausnahme darstellen, die im Rahmen der vorliegenden Dissertation allerdings nicht weiter thematisiert wird.

⁹ Hierbei ist es allerdings zu hinterfragen, ob in diesen Fällen wirklich der Technologie an sich vertraut wird, da deren Aktionen determiniert sind und diese über keine mentalen Zustände verfügt (vgl. Kapitel 3.4.5).

Notwendigkeit (*trust necessity*), die entsprechenden kognitiven Fähigkeiten (*cognitive ability*), um die Vertrauenswürdigkeit einzuschätzen, und die Disposition zum Vertrauen (*trust propensity*) relevant (Juvina et al., 2019).

Vertrauen wird zumeist als graduelle Größe verstanden. Im Gegensatz zu einer dichotomen Unterscheidung zwischen *Vertrauen* und *kein Vertrauen* bewegt sich das Maß an Vertrauen demnach auf einem Kontinuum, dessen Endpunkte, d. h. *kein Vertrauen* und *maximales Vertrauen*, allerdings kaum zu bestimmen sind. Vertrauen unterliegt dabei einem dynamischen Veränderungsprozess und wird aktualisiert durch neue Erfahrungswerte, Erkenntnisse und daraus abgeleitete abweichende Erwartungen in Hinblick auf das Verhalten der Vertrauensnehmerin oder des Vertrauensnehmers. Dabei wirken sich als negativ wahrgenommene Ereignisse, also Vertrauensverletzungen, stärker auf das Vertrauensniveau aus als positive Ereignisse, zumal sie häufig deutlich sichtbare negative Konsequenzen nach sich ziehen (Flook et al., 2019; J. D. Lee & See, 2004, S. 72). Der Zusammenhang zwischen Positivereignissen und dem Vertrauensniveau ist demnach keineswegs proportional. Vertrauen ist vielmehr schwer zu gewinnen, aber einfach zu verlieren (Flook et al., 2019). Insbesondere zu Beginn des Prozesses der Vertrauensbildung, wenn nur wenige Erfahrungswerte vorliegen, ist das Vertrauen vulnerabel (Desai, Kaniarasu, Medvedev, Steinfeld & Yanco, 2013). Dennoch führt nicht jede Negativerfahrung zwangsläufig zu einem Entzug des Vertrauens. Es ist im Gegenteil durchaus möglich, dass der vertrauensnehmenden Instanz ein gewisser Kredit eingeräumt wird, „in dessen Rahmen auch ungünstige Erfahrungen zurechtinterpretiert oder absorbiert werden können“ (Luhmann, 1968/2014, S. 37) und aus Vereinfachungsgründen das Vertrauen so lange aufrechterhalten wird, bis eine gewisse Schwelle an Negativerfahrungen überschritten ist. Dabei können besonders symbolische vertrauensverletzende Aktionen eine überproportionale Bedeutung einnehmen (Luhmann, 1968/2014, S. 37).

3.4.2 Vertrauen als mentaler Zustand

Auch die Festlegung, um welche Art von mentalem Zustand es sich beim Vertrauen handelt, erweist sich als herausfordernd. Vertrauen lässt sich als Überzeugung (*belief*), Einstellung (*attitude*), Absicht (*intention*) oder als Handlung selbst (*behavior*) konzeptionieren. Diese Bestimmung hat u. a. erhebliche erkenntnistheoretische Konsequenzen, denn Einstellungen lassen sich nur erfragen, wohingegen Handlungen für Dritte beobachtbar sind. Die Unterscheidungen und Abhängigkeiten zwischen diesen Zuständen lassen sich aus der *Theorie des überlegten Handelns* (*theory of reasoned action*, TRA; Ajzen & Fishbein, 1980; Fishbein & Ajzen, 1975) ableiten, auf die einflussreiche Vertrauensmodelle Bezug nehmen (vgl. Kapitel 3.4.7). Die TRA betrachtet Handlungen als Konsequenzen von Handlungsabsichten, die sich aus einem Zusammenspiel von bestimmten Überzeugungen und Einstellungen ergeben, wie Abbildung 3.4 zeigt.

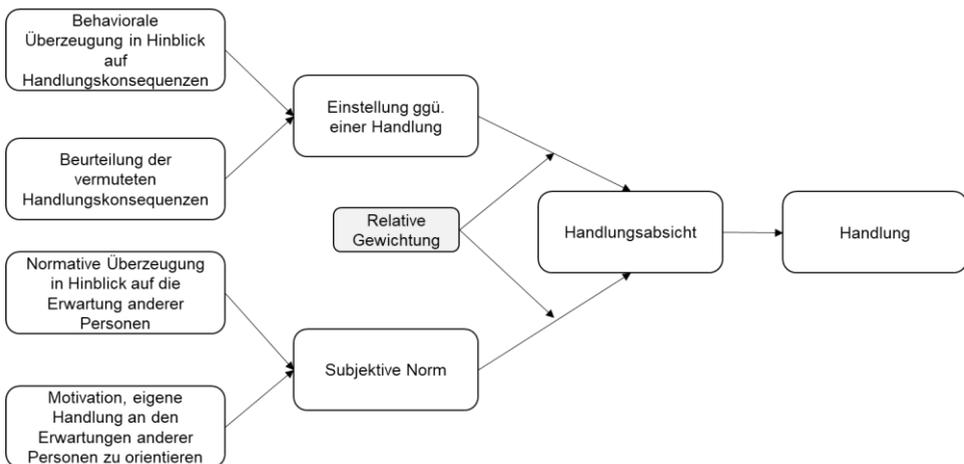


Abbildung 3.4: Theorie des überlegten Handelns, leicht adaptiert aus Fishbein und Ajzen (1975, S. 16) und Ajzen und Fishbein (1980, S. 8).

Die persönliche Einstellung gegenüber einer Handlung bildet sich einerseits aus der behavioralen Überzeugung, mit einem gewissen Verhalten ein Verhaltensergebnis erzielen zu können, und andererseits aus einer positiven Bewertung dieses erwarteten Verhaltensergebnisses. Hierbei lassen sich die Stärke der Überzeugung im Sinne einer geschätzten Auftretenswahrscheinlichkeit des Verhaltensergebnisses sowie die Stärke der positiven Bewertung in der Tradition von Erwartungswert-Theorien miteinander multiplizieren, um die Stärke der Einstellung zu ermitteln (Winter, 2013, S. 173ff.). Demnach sind die jeweiligen Ausprägungen der mentalen Zustände nicht als dichotom, sondern als graduell, d. h. in ihrer Intensität variierend, angelegt. Der persönlichen Einstellung steht die sogenannte subjektive Norm gegenüber. Diese umschreibt, wie andere Individuen oder Gruppen die geplante Handlung vermutlich beurteilen. Außerdem fließt die subjektive Motivation des oder der Agierenden mit ein, sein oder ihr Handeln an den Meinungen von Dritten zu orientieren. Die subjektive Norm kann somit als eine Form des sozialen Drucks angesehen werden (Ajzen & Fishbein, 1980, S. 7). Auf Basis der Einflüsse dieser beiden zugrundeliegenden Determinanten entwickeln Personen eine Absicht, entweder eine Handlung zu verrichten oder dies zu unterlassen (Ajzen & Fishbein, 1980, S. 6). Aus einer etwaigen Handlungsabsicht folgt im Regelfall die tatsächliche Handlung (Ajzen & Fishbein, 1980, S. 5).

Während sich die TRA nur auf überlegtes Handeln anwenden lässt, umfasst die *Theorie des geplanten Handelns* (theory of planned behavior; TPB; Ajzen, 1985), die als Erweiterung der TRA gilt, auch routinemäßige Handlungen, die häufig nicht unter der bewussten willentlichen Kontrolle der oder des Agierenden stehen. Die TPB enthält zusätzlich zur TRA die Komponente der wahrgenommenen Verhaltenskontrolle (*control beliefs*). Diese

beschreibt, inwiefern die Person die tatsächliche Durchführung einer Handlung für möglich hält. Damit weist die wahrgenommene Verhaltenskontrolle einen engen Bezug zur empfundenen Selbstwirksamkeit auf und bezieht neben internen Faktoren wie mangelnde Kompetenzen auch externe Faktoren wie mangelnde Ressourcen (z. B. zeitliche, finanzielle, infrastrukturelle) ein (Ajzen, 1985).

3.4.3 Vertrauensdefinition im MRI-Kontext

Trotz einer langen Historie an Versuchen, das Konstrukt des Vertrauens genauer zu bestimmen, existiert bis dato keine weit verbreitete, allgemein anerkannte Definition für Vertrauen (Schaefer, 2013, S. 19). Die Folge ist eine Vielzahl verschiedener Ansätze zur Konzeptualisierung von Vertrauen (J. D. Lee & See, 2004, S. 52), die sich gemäß TRA unterschiedlichen mentalen Zuständen zuordnen lassen. Tabelle 12 listet einige einflussreiche Definitionen mit den jeweiligen Zuordnungen auf.

Tabelle 12: Zuordnung unterschiedlicher Vertrauensdefinitionen zu den mentalen Zuständen gemäß der TRA, größtenteils angelehnt an J. D. Lee und See (2004).

Zustand gemäß TRA	Definition (dt., eigene Übersetzung)	Definition (engl., wörtlich)	Autor:innen
Überzeugung	Erwartung einer Person, sich auf das Wort, das Versprechen oder die schriftliche Kommunikation einer anderen Person verlassen zu können.	Expectancy held by an individual that the word, promise or written communication of another can be relied upon.	Rotter (1967, S. 651)
Einstellung	Einstellung, dass ein:e Akteur:in einem Individuum beim Erreichen von dessen Zielen in einer von Unsicherheit und Vulnerabilität gekennzeichneten Situation helfen wird.	The attitude that an agent will help achieve an individual's goals in a situation characterized by uncertainty and vulnerability.	J. D. Lee und See (2004, S. 54)
Handlungsabsicht	Bereitschaft einer Partei, sich gegenüber den Aktionen einer anderen Partei verwundbar zu machen, auf Basis der Erwartung, dass die andere Partei eine	Willingness of a party to be vulnerable to the actions of another party based on the expectation that the other will perform a particular action important to the trustor,	Mayer et al. (1995, S. 712)

Zustand gemäß TRA	Definition (dt., eigene Übersetzung)	Definition (engl., wörtlich)	Autor:innen
	bestimmte, für die Vertrauensgeberin oder den Vertrauensgeber relevante Aktion ausführen wird, ungeachtet der Fähigkeit, die andere Partei zu überwachen oder zu kontrollieren.	irrespective of the ability to monitor or control that party.	
Handlung	Zustand wahrgenommener Verwundbarkeit und wahrgenommenen Risikos, der sich aus einer Unsicherheit bezüglich der Motive, Absichten und perspektivischen Handlungen eines Gegenübers ergibt, von denen die Person abhängt. Der Akt des Sich-Verlassens darauf, dass Aktionen, die dem eigenen Wohlergehen schaden, von einflussreichen Anderen unterlassen werden.	A state of perceived vulnerability or risk that is derived from an individual's uncertainty regarding the motives, intentions, and perspective actions of others on whom they depend. The reliance by an agent that actions prejudicial to their wellbeing will not be undertaken by influential others.	Kramer (1999, S. 571) Hancock, Billings und Schaefer (2011, S. 24)

Die Unterscheidungen sind jeweils nicht als trennscharf anzusehen. So differenzieren J. D. Lee und See (2004) in ihrer viel beachteten Meta-Analyse zum Vertrauen in Automationstechnologie die Definitionen bspw. nicht zwischen Überzeugungen und Einstellungen. Eine klare Trennung nehmen sie allerdings anhand des Begriffs des Sich-auf-etwas-Verlassens (*reliance*) vor. Sich auf etwas zu verlassen, erachten sie als Ausdruck verhaltensbasierten Vertrauens: „Im Kontext des Vertrauens und Sich-Verlassens ist das Vertrauen eine Einstellung und das Sich-auf-etwas-Verlassen ist eine Verhaltensweise“ (*in the context of trust and reliance, trust is an attitude and reliance is a behavior*) (J. D. Lee & See, 2004, S. 53). Die meisten Arbeiten zum Vertrauen im MRI-Kontext rekurrieren auf o. g. Vertrauensdefinition nach J. D. Lee und See (2004), die einst allgemeiner für das Vertrauen in Automatisierungstechnologie formuliert wurde (Spain, Bustamante & Bliss, 2008). Zunächst erweckt diese Definition den Anschein, als konzeptualisiere sie Vertrauen als bloße Überzeugung in Hinblick auf die Zukunft und als lasse sie damit die normative Komponente einer Einstellung vermissen. Allerdings impliziert die Formulierung des

Helpens bei der Erreichung eines *Ziels* eine positive Evaluation, sodass die Definition somit die Kriterien für den mentalen Zustand einer Einstellung erfüllt. Der verwendete Begriff des Akteurs bzw. der Akteurin kann hierbei sowohl Maschinen als auch Personen umfassen. Bemerkenswert hierbei ist, dass die Autor:innen diese Definition selbst als „simple (...) Basis-Definition“ (J. D. Lee & See, 2004, S. 54) charakterisieren und darauf verweisen, dass die bis dato anerkannteste Definition von Mayer et al. (1995) stammt.

Im Rahmen der vorliegenden Dissertation ist zu prüfen, inwiefern sich die Definition nach J. D. Lee und See (2004) sinnvoll auf das Mensch-Roboter-Vertrauen am industriellen Arbeitsplatz anwenden lässt. Gemäß der Definition von J. D. Lee und See (2004) ließe sich das Zustandekommen von Vertrauen in einen Roboter am industriellen Arbeitsplatz unter Rückgriff auf die TRA folgendermaßen exemplarisch ausbuchstabieren: Die vertrauensgebende Person erwartet, dass ein Roboter in der Lage ist, Werkstücke in die richtige Position zur Weiterverarbeitung zu bringen (*Überzeugung*) und bewertet dies als hilfreich in Hinblick auf die Erreichung ihrer Ziele (*subjektive Bewertung*). Die Person entwickelt also eine positive Einstellung (*Einstellung*) dem Roboter gegenüber. Auf Basis dieses einstellungsbasierten Vertrauens ist sie bereit, trotz der Unsicherheit über die tatsächlichen Aktionen des Roboters auf dessen permanente Kontrolle zu verzichten. Da die Person davon ausgeht, dass andere Mitarbeitende dem Roboter ebenfalls positiv gegenüber eingestellt sind oder zumindest seine Einstellung gegenüber dem Roboter goutieren (*subjektive Norm*), erwächst daraus eine Absicht, den zur Verfügung gestellten Roboter tatsächlich zu verwenden. Vorbehaltlich weiterer externer Faktoren führt dies dazu, dass die Person den Roboter tatsächlich benutzt. Im Zusammenhang mit Automatisierung am Arbeitsplatz ist allerdings zu bedenken, dass Verhaltensabsichten oder tatsächliches Verhalten womöglich durch äußere Nutzungszwänge herbeigeführt werden, da Mitarbeitende keine vollumfängliche Entscheidungsfreiheit hinsichtlich der Nutzung einer Automatisierungslösung besitzen (Oubari et al., 2018, S. 560f.). Demnach drückt das Verhalten der Mitarbeitenden in Bezug auf eine Technologie nicht zwangsläufig deren Einstellung gegenüber dieser Technologie aus, sondern kann bspw. deren Bereitschaft indizieren, Anordnungen zu befolgen.

Ferner verweist die Vertrauensdefinition nach J. D. Lee und See (2004) auf eine Situation der Vulnerabilität, die das Vertrauen erfordert, und auf die Behilflichkeit der vertrauensnehmenden Instanz bei der Zielerreichung. Entsprechend sind die typischen Vulnerabilitäten, denen sich Mitarbeitende am industriellen Arbeitsplatz ausgesetzt sehen, und die Ziele, die sie an demselben verfolgen, zu spezifizieren. Zu den Vulnerabilitäten könnten körperliche Verletzungen durch den Roboter, ein selbstwertbedrohendes Gefühl der Überforderung, schlechtere Arbeitsleistungen im Zusammenspiel mit dem Roboter oder Angst vor technologischer Ersetzung und Arbeitsplatzverlust gehören. Zu den typischen Zielen wäre etwa die erfolgreiche Bewältigung der Arbeitsaufgaben, die körperliche Unversehrtheit, Spaß an der Arbeit und eine langfristige Sicherung der Beschäftigung zu zählen. Diese Aufzählung verdeutlicht, dass die Erreichung dieser Ziele allenfalls zu einem geringen

Anteil durch die Ausgestaltung der Technologie, sondern vielmehr durch die Vertrauenswürdigkeit der an der Technologieeinführung und -nutzung beteiligten Akteur:innen und Institutionen beeinflusst wird (vgl. Kapitel 3.4.5). Insofern wird die vorliegende Definition in dieser Dissertation nicht so verstanden, dass die Technologie als vertrauensnehmende Instanz auftritt, wenngleich die aktive Formulierung, dass ein:e Akteur:in bei der Zielerreichung helfen wird (*will help*), eine gewisse Autonomie nahelegt. Um diese Konnotation zu vermeiden, ließe sich in einer Passivkonstruktion davon sprechen, dass ein:e Akteur:in bei der Zielerreichung hilfreich sein wird (*will be helpful*), sei es in Form eines simplen Werkzeugs oder einer elaborierteren interaktiven Technologie.

3.4.4 Kognitives und affektives Vertrauen

In der Sozialpsychologie gilt es als *common sense*, dass Einstellungen grundsätzlich eine kognitive sowie eine affektive Komponente besitzen (Trafimow & Sheeran, 2004). Dies gilt ebenfalls für als Einstellung konzeptioniertes Vertrauen (B. French et al., 2018, S. 26). Dabei lässt sich kognitives Vertrauen als derjenige Anteil verstehen, der auf Basis von (vermeintlichem) Wissen über die Fähigkeiten und die Zuverlässigkeit einer vertrauensnehmenden Instanz entsteht, wohingegen affektives Vertrauen aus einer spontanen Reaktion auf das beobachtete Verhalten derselben resultiert (Gompei & Umemuro, 2018, S. 46). Kognitives Vertrauen bezieht explizit frühere Erfahrungen mit dem Gegenüber und dessen Reputation oder Vertrauenswürdigkeit mit ein, wohingegen affektives Vertrauen auf der emotionalen Verbindung zum Gegenüber basiert und dabei auch diesem zugeschriebene Eigenschaften und Motive einbezieht (Bernotat, Eyssel & Sachse, 2019). Die Relevanz der affektiven Vertrauenskomponenten bringen J. D. Lee und See (2004, S. 63) in Anlehnung an Fine und Holyfield (1996) mit der Aussage auf den Punkt, dass „Menschen nicht nur über Vertrauen nachdenken, sondern Vertrauen auch verspüren“ (*people not only think about trust, they also feel it*). Die Autor:innen betrachten Vertrauen als eine letztlich affektive Reaktion, die zwar auch mit analytischen und analogischen Vertrauensbildungsprozessen in Wechselwirkung steht, bei der die affektiven Prozesse allerdings den stärksten Einfluss ausüben (J. D. Lee & See, 2004, S. 61). So versagen gewisse Personen anderen Personen das Vertrauen mitunter aufgrund unbehaglicher Gefühle, die sich kaum rationalisieren lassen (Hoff & Bashir, 2015). Dies deckt sich mit Befunden aus der Akzeptanzforschung, die ebenfalls den Einfluss subtiler, mitunter leicht beeinflussbarer subjektiver Bewertungen und Gefühlslagen hervorheben (Meißner & Trübswetter, 2018, S. 227; Onnasch et al., 2019, S. 38). Trotz der deutlichen Hinweise auf die Relevanz affektiver Prozesse besteht insgesamt noch keine Gewissheit darüber, zu welchen Anteilen und in welchen Kontexten kognitive und affektive Komponenten die Vertrauensbildung letztlich beeinflussen (B. French et al., 2018, S. 26). Obgleich die TRA grundsätzlich als empirisch gut validiert gilt (Winter, 2013, S. 177f.), unterstreicht die Relevanz affektiver Komponenten bei der Vertrauensbildung die begrenzte Anwendbarkeit der TRA in diesem Kontext. Die

Attraktivität der TRA liegt u. a. in ihrer Sparsamkeit begründet, die gleichzeitig insofern einen Nachteil darstellt, als bspw. antizipierte Emotionen während der Handlungsausführung oder bei Eintritt der vermuteten Handlungskonsequenzen in der Theorie außer Acht gelassen werden, obwohl einige Studien deren Relevanz belegen konnten (D. P. French et al., 2005).

Im zeitlichen Verlauf einer MRI ist davon auszugehen, dass der Einfluss von kognitiven und affektiven Komponenten variiert. Affektive Reaktionen treten v. a. im direkten Kontakt mit der Technologie in Erscheinung, wohingegen kognitives Vertrauen auch im Vorfeld desselben auf Basis rationaler Abwägungen entstehen kann. Affektives Vertrauen lässt sich daher allenfalls dynamisch innerhalb realer Interaktionen untersuchen. Da solche experimentellen Settings aufwändig und komplex sind, wurden affektive Vertrauensanteile bisher vergleichsweise wenig untersucht (Gompei & Umemuro, 2018, S. 46; Lewicki et al., 2006). Die stärkste affektive Reaktion wird am Beginn einer Interaktion vermutet, die sich dann im Laufe der Interaktionszeit stabilisiert (Gompei & Umemuro, 2018, S. 52). Während J. D. Lee und See (2004, S. 61) Vertrauen letztlich als affektive Reaktion begreifen, schlussfolgern Lewis et al. (2018, S. 149), dass im Kontext des Mensch-Technik-Vertrauens die kognitiven Komponenten eher dominieren, da hier Überlegungen zur Zuverlässigkeit im Sinne der Übereinstimmung von erwarteten und tatsächlichen Aktionen einer Maschine im Vordergrund stehen. Ob sich dies so pauschal und insbesondere in Bezug auf stark interaktive Technologie konstatieren lässt, ist allerdings fraglich. Gerade Vermenschlichungsprozesse adressieren u. a. emotionale Aspekte wie die Zuneigung zum Roboter und wirken sich somit merklich auf affektives Vertrauen aus (Cameron et al., 2016). Einen Schwachpunkt vieler MRI-Studien stellt in diesem Zusammenhang die fehlende explizite Differenzierung zwischen affektiven und kognitiven Vertrauensanteilen dar, die im Regelfall weder konzeptionell noch im Rahmen empirischer Messungen vorgenommen wird.

3.4.5 Mögliche Adressat:innen des Vertrauens im Mensch-Technik-Kontext

Die vorgenannte Bestimmung des Vertrauens als soziales Phänomen setzt die Beteiligung zweier Personen voraus und eröffnet somit die Fragestellung nach der Übertragbarkeit dieses Vertrauenskonzepts auf Mensch-Technik-Vertrauen im Allgemeinen und Mensch-Roboter-Vertrauen im Speziellen, das im Kontext der sozialen Robotik in den vergangenen Jahren zunehmend erforscht wurde (Powell & Michael, 2019). Während der Begriff in seiner alltagssprachlichen Verwendung impliziert, dass das Vertrauensverhältnis zwischen einem Menschen und einem Roboter entsteht, erscheint die tatsächliche Sachlage komplexer. Denn wie in Kapitel 3.4.1 bereits erwähnt, gilt als Voraussetzung für Vertrauen, dass beide beteiligten Parteien in ihren inneren Vorgängen und äußeren Handlungen nicht vollständig determiniert sind, sich also frei für oder gegen eine Handlung entscheiden können

und damit eine Situation der Unsicherheit für das jeweilige Gegenüber schaffen. Zwar scheint das bei Robotern nicht gegeben, allerdings stellt sich die Frage, ob diese dennoch als vertrauensnehmende Instanz auftreten können, insofern sie Menschen als autonome Akteur:innen *erscheinen*. Aus einer solchen konstruktivistischen Perspektive auf Vertrauen, die den Stellenwert von Zuschreibungen betont, vertritt Sumpf (2019, S. 7) die Auffassung, dass jedwedes Vertrauen letztlich als Systemvertrauen betrachtet werden kann und damit verschiedenartige Adressat:innen einschließt. Damit lassen sich auch Technologien als vertrauensnehmende Instanzen einbeziehen. Das gilt insbesondere für technische Artefakte, die Personen mental als quasi-soziales Objekt repräsentieren und zu denen sie eine soziale Beziehung konstruieren. In dem Maße, wie die Komplexität und Autonomie der Technologien zunimmt, versetzen sich Menschen mehr und mehr in eine Art Beziehung zu der Technologie, die sich durch eine wechselseitige Interaktion anstatt durch ein Verhältnis des (einseitigen) Sich-Verlassens auf die Funktionsfähigkeit der Maschine charakterisieren lässt (Atkinson et al., 2012). Wenngleich das Vertrauen und das Sich-Verlassen somit wechselseitig zu sein scheint, bleibt es aufgrund der Unbelebtheit technischer Artefakte letztlich doch ein einseitiger Vorgang (Hancock et al., 2020). Jedenfalls muss die Technologie im Vertrauensbildungsprozess nicht notwendigerweise involviert sein. Vielmehr entwickelt sich der Prozess der Vertrauensbildung im Verhältnis zwischen dem Menschen und den Technikentwickler:innen, den Regulationssystemen usw. (Sumpf, 2019, S. 112f.).

Damit bleibt die komplexe Frage bestehen, wem oder was letztlich Vertrauen entgegengbracht wird. Im Kontext der industriellen MRI lässt sich annehmen, dass die Mitarbeitenden im Vorfeld der Interaktion bspw. darauf vertrauen, dass institutionell abgesichert ist, dass von der Maschine keine Gefahr für ihr Leib und Leben ausgeht, dass der Hersteller den Roboter derart programmiert hat, dass er einfach zu bedienen ist, und dass die Geschäftsleitung mit der Einführung des Roboters nicht die Ersetzung oder Degradierung der Mitarbeitenden beabsichtigt. Dieser organisationale Kontext, der auch von Gerüchten zur Vertrauenswürdigkeit der Technologie geprägt wird, spielt im Vorfeld einer Robotereinführung eine erhebliche Rolle für das Vertrauen (J. D. Lee & See, 2004, S. 56). Erst während der späteren tatsächlichen Interaktion mit dem Roboter rückt dessen wahrgenommene Verlässlichkeit in den Fokus und wird zum (mit-)bestimmenden Faktor für das Vertrauen.

3.4.6 Auswirkungen auf Technologienutzung

Gemeinhin wird Vertrauen als notwendiges Kriterium für erfolgreiche Mensch-Mensch- und Mensch-Roboter-Interaktionen angesehen, wobei sich dieser Erfolg wiederum auf höchst unterschiedliche Kriterien wie performative Kenngrößen oder die Zufriedenheit der menschlichen Teammitglieder beziehen kann (Broadbent, 2017; Charalambous et al., 2013, 2016b; Ososky, 2013, S. 27; Phillips et al., 2011; Strohkorb Sebo, Traeger, Jung &

Scassellati, 2018; You & Robert, 2019). Die Relevanz von Vertrauen im MRI-Kontext speist sich in erster Linie aus der Annahme, dass (einstellungsbezogenes) Vertrauen die Wahrscheinlichkeit erhöht, dass die vertrauende Person die Technologie nutzt (Sanders, Kaplan, Koch, Schwartz & Hancock, 2019) und sich dabei tatsächlich auf sie verlässt (J. D. Lee & See, 2004). Es stehen also sowohl die Nutzungsentscheidung selbst als auch die Art und Weise der Nutzung zur Disposition. Vertrauen stellt dabei zwar ein begünstigendes, allerdings kein hinreichendes Kriterium für eine entsprechende Technologienutzung dar, da weitere Überlegungen und externe Faktoren in den Prozess einfließen (J. D. Lee & See, 2004). Demnach führt weder Vertrauen zwangsläufig zu einer Nutzungsentscheidung noch ist jede Nichtbenutzung einer Technologie zwangsläufig durch mangelndes Vertrauen begründet (Sanders et al., 2019). So könnten bspw. zu geringe (wahrgenommene) praktische Vorteile im Vergleich zum Aufwand der Nutzung oder eine mangelnde Notwendigkeit, sich tatsächlich auf eine Technologie zu verlassen¹⁰, ursächlich für eine Nichtbenutzung sein (Gaudiello, Zibetti, Lefort, Chetouani & Ivaldi, 2016; J. D. Lee & See, 2004). Umgekehrt kann eine Nutzungsentscheidung ohne Vertrauen durch andere Faktoren wie bspw. durch das Vorhandensein eines Gefühls der Verbindlichkeit (*sense of commitment*) oder schlicht durch äußere Anordnungen in einem verpflichtenden Nutzungskontext motiviert sein (Michael & Salice, 2017; Oubari et al., 2018; Powell & Michael, 2019). Devitt (2018, S. 172) weist in diesem Zusammenhang darauf hin, dass sich Menschen im Alltag häufig auf Dinge verlassen, ohne ihnen zu vertrauen, z. B. wenn sie Nutzungsbedingungen von Software bestätigen, ohne diese überhaupt gelesen zu haben. Hierbei ist allerdings fraglich, ob der Verzicht auf das Lesen der Nutzungsbedingungen nicht bereits als Konsequenz vorhandenen Vertrauens gelten kann. Die tatsächlichen Beweggründe sind nur schwer zu ermitteln, zumal Menschen dazu zu neigen scheinen, den Begriff des Vertrauens eher in Bezug zu Menschen anstatt in Bezug zu Robotern als Begründung anzuführen (Sanders et al., 2019).

Obgleich die Zusammenhänge zwischen dem Vertrauen und dem Sich-auf-etwas-Verlassen nicht eindeutig sind, lässt sich als Konsens festhalten, dass das Vorhandensein eines gewissen Vertrauens die Wahrscheinlichkeit einer Nichtbenutzung (*disuse*) einer Technologie verringert¹¹, sofern im jeweiligen Nutzungskontext überhaupt eine freie Entscheidung möglich ist. Ein angemessenes Maß an Vertrauen vermindert ferner die Gefahr, dass es zu einer unsachgemäßen Benutzung (*misuse*) aufgrund einer zu niedrigen oder zu hohen

¹⁰ Diese Argumentation von Gaudiello, Zibetti, Lefort, Chetouani und Ivaldi (2016), dass aus (einstellungsbezogenem) Vertrauen kein (handlungsbezogenes) Sich-Verlassen auf die vertrauensnehmende Instanz folgt, sofern es keine Notwendigkeit dazu gibt, mag zwar zutreffen, klammert aber den Umstand aus, dass das Konzept des Vertrauens ohnehin nur in Situationen der Unsicherheit relevant ist, also in diesem Szenario qua Definition kein Vertrauen, sondern lediglich eine Erwartungshaltung vorliegt.

¹¹ Die Entscheidung zur Nichtbenutzung wird hier mit einem Mangel an Vertrauen in Verbindung gebracht. Alternativ ließe sich Nichtbenutzung als Folge von Misstrauen anstatt als Folge mangelnden Vertrauens verstehen (vgl. Kapitel 3.4.10).

Erwartungshaltung (*over-/underestimation*), eines zu niedrigen oder zu hohen Vertrauensniveaus (*over-/undertrust*) oder eines zu niedrigen oder zu hohen Ausmaßes des Sich-Verlassens auf die Technologie (*over-/underreliance*) kommt (Hancock et al., 2020; J. D. Lee & See, 2004; Lewis et al., 2018; Parasuraman & Riley, 1997). Je mehr sich eine Person auf eine Technologie verlässt, desto mehr wird sie bereit sein, auf Kontrolle zu verzichten. Ein kompletter Verzicht auf Kontrolle wäre in der Praxis allerdings ebenso wenig wünschenswert wie eine permanente Kontrolle. Letztere würde die zur Verfügung stehende Zeit zur Verrichtung eigener Aufgaben verringern, wohingegen ein kompletter Kontrollverzicht bspw. dazu führen könnte, dass etwaige Fehlfunktionen eines Roboters nicht erkannt werden (Hancock et al., 2020; Lewis et al., 2018). Wenn die Erwartungen der Anwender:innen in Bezug auf die Technik überzogen sind oder aus anderen Gründen nicht erfüllt werden, kann dies das Vertrauen unterminieren, selbst wenn die Technologie wie beabsichtigt funktioniert (J. D. Lee & See, 2004, S. 72; Powell & Michael, 2019). Eine konservativere Erwartungshaltung und ein geringeres Vertrauensniveau gegenüber der Technologie können sich daher als vergleichsweise vorteilhaft erweisen (Yang, Unhelkar, Li & Shah, 2017, S. 414). Obwohl diese Abwägungen verdeutlichen, dass zu geringe und zu hohe Vertrauensniveaus gleichermaßen negative Konsequenzen haben können, wird häufig ein Zustand geringen Vertrauens einseitig als Mangel aufgefasst, der eine möglichst schnelle Behebung erfordert. In dieser Vorstellung erscheint das Herstellen von Vertrauen *per se* als gut (Sumpf, 2019, S. 90). Dabei wird häufig übersehen, dass es durchaus wünschenswert ist, nicht vertrauenswürdigen Entitäten kein Vertrauen zu schenken. Wenn Forschende und Praktiker:innen für eine Steigerung des Vertrauens plädieren, sollte daher stets die tatsächliche Vertrauenswürdigkeit berücksichtigt werden (Hardin, 2006, S. 1f./67).

3.4.7 Vertrauensmodelle

In diesem Kapitel werden in den ersten drei Unterkapiteln wichtige Vertrauensmodelle dargestellt, die im vierten Kapitel gemeinsam mit weiteren Differenzierungen im Bereich des Vertrauens, die in den vorigen Kapiteln erläutert wurden, zu einem eigenen integrierten Vertrauensmodell zusammengeführt werden. An dieser Stelle sei darauf hingewiesen, dass weitere Vertrauensmodelle existieren (z. B. de Visser et al., 2020), die im Rahmen der vorliegenden Dissertation nicht näher thematisiert werden.

3.4.7.1 Basis-Modell nach Lee & See (2004)

Die Forschung zum Vertrauen in Automatisierung wurde maßgeblich durch die Literaturanalyse von J. D. Lee und See (2004) und dem daraus resultierenden Modell zur Dynamik des Vertrauens in Automatisierung geprägt (B. French et al., 2018, S. 32). Wie die Technologieakzeptanzmodelle (Davis, 1985) beruht das Modell auf der TRA und differenziert verschiedene Stadien der Vertrauensentwicklung. Wie Abbildung 3.5 zeigt, entsteht ausgehend von Erwartungen aufgrund assimilierter Informationen das einstellungsbasierte

Vertrauen, dann eine Nutzungsintention und schließlich die Bereitschaft, eine Technologie tatsächlich zu nutzen und sich auf diese zu verlassen. Dabei sind die Informationen, die zur Ableitung von Erwartungen herangezogen werden, nicht als Wissen über die Technologie zu verstehen, sondern werden u. a. über Dritte in Form von Erzählungen oder über eine bestehende Reputation vermittelt. Entscheidender Aspekt des Modells ist der geschlossene Kreislauf in Form einer „Rückkopplungsschleife“ (Luhmann, 1968/2014, S. 37), bei dem die Erfahrungen aus der realen Nutzung der Technologie stets zur Aktualisierung des bestehenden Vertrauensniveaus herangezogen werden. Diese Dynamik des Modells erlaubt eine regelmäßige, iterative Überprüfung der Angemessenheit des Vertrauens, die sich im Modell als die Beziehung zwischen dem aktuellen Grad des Vertrauens und der damit verbundenen Erwartungshaltung sowie den beobachteten Fähigkeiten der Automatisierungstechnologie darstellt.

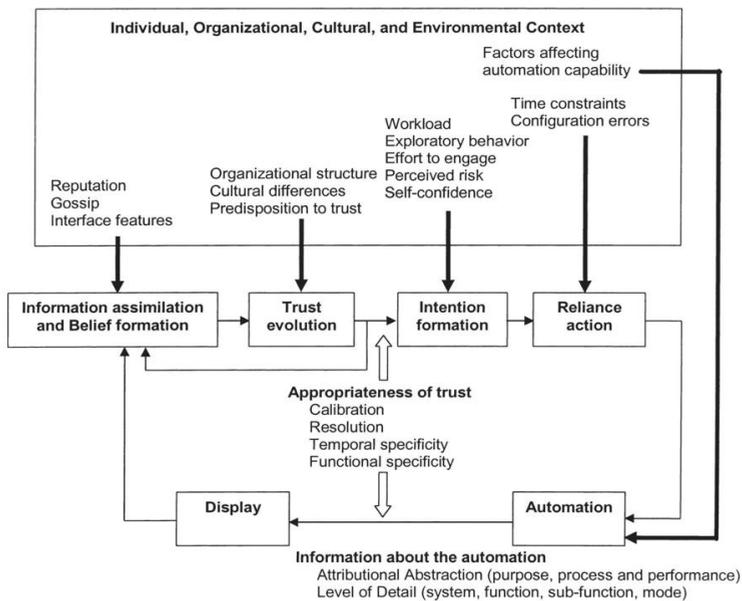


Abbildung 3.5: Modell zum Vertrauen in Automatisierung von J. D. Lee und See (2004, S. 68). Copyright © 2004 Human Factors and Ergonomics Society. Nachdruck mit Genehmigung von SAGE Publications.

Im Verlauf der Interaktion wird das Vertrauen abhängiger von den realen Beobachtungen. Die ungesicherte Reputation wird zunehmend durch eigene, erfahrungsbasierte Beurteilungen aktualisiert und ersetzt. Gemäß J. D. Lee und See (2004) stellt die Entscheidung, sich auf die Technologie zu verlassen, eine notwendige Bedingung dafür dar, dass positive Nutzungserfahrungen entstehen und das Vertrauen wachsen kann. Fällt das Vertrauen zunächst zu gering aus, sodass die vertrauende Person nicht dazu bereit ist, sich auf eine Technologie

zu verlassen, bleiben ihr daher entscheidende Möglichkeiten für ein substanzielles Vertrauenswachstum verwehrt (Muir & Moray, 1996).

Trotz der häufigen Rezeption des Modells nach Lee & See (2004) und der dazugehörigen Vertrauensdefinition wurde das Modell kaum konkret in empirischen Untersuchungen angewandt und validiert. Insbesondere die Unterscheidungen zwischen Überzeugung, Einstellung und Absicht sind in experimentellen Settings schwer vorzunehmen und zu operationalisieren. Daher sprechen viele Studien simplifizierend von einer Untersuchung des Vertrauens, ohne dass die Ebene spezifiziert wird, auf der sich das untersuchte Vertrauen verorten lässt (B. French et al., 2018).

3.4.7.2 Drei-Schichten-Modell nach Hoff & Bashir (2015)

Das *Drei-Schichten-Modell* von Hoff und Bashir (2015) basiert auf einer Meta-Analyse von 101 Veröffentlichungen. Die Autor:innen fokussierten sich dabei ausdrücklich auf Vertrauen in Automation, bezogen aber auch Studien zum Vertrauen in soziale Roboter mit ein, um empirische Lücken in der bestehenden Forschungsliteratur zu kompensieren. Das Drei-Schichten-Modell differenziert drei wesentliche Ebenen der Vertrauensbildung, nämlich dispositionale, situative und vermittelte Faktoren¹², die ursprünglich von Marsh und Dibben (2003) formuliert wurden. Hoff und Bashir (2015) identifizierten als Ergebnis ihrer Meta-Analyse drei Ursachen für unterschiedliche Vertrauensniveaus, nämlich Eigenschaften der vertrauenden Person, der Umgebung und des technischen Systems selbst. Wie Tabelle 13 verdeutlicht, lassen sich diese zu den drei Vertrauensfaktoren nach Marsh und Dibben (2003) in Beziehung setzen, wenngleich es sich dabei nicht um eindeutige Eins-zu-eins-Zuordnung handelt.

Tabelle 13: Einflussfaktoren auf Vertrauen im Drei-Schichten-Modell.

Vertrauensfaktoren nach Marsh und Dibben (2003)	Vertrauensfaktoren nach Hoff und Bashir (2015)	Beschreibung
Dispositional	Menschliche:r Operator:in	Zeitstabile, persönlichkeitsbezogene Tendenz zum Vertrauen

¹² Der englische Ausdruck *learned trust* aus der Primärliteratur wird fortan mit *vermitteltem Vertrauen* übersetzt, um insbesondere den widersprüchlich anmutenden Ausdruck des initialen *erlernten* Vertrauens zu vermeiden und zu betonen, dass diese Art von Vertrauen nicht nur durch eigene Erfahrungen, sondern auch durch die Vermittlung Dritter entsteht. Die Übersetzung als *vermitteltes Vertrauen* schließt ein, dass diese Art von Technikvertrauen durch gedankliche Übertragungen von Erfahrungen mit als ähnlich empfundenen Entitäten, durch sprachliche Frames oder durch Aussagen Dritter vermittelt wird (initiales Vertrauen *a priori*) oder auf konkreten eigenen Erfahrungen mit der jeweiligen Technologie (dynamisches Vertrauen *a posteriori*) basieren kann.

Vertrauensfaktoren nach Marsh und Dibben (2003)	Vertrauensfaktoren nach Hoff und Bashir (2015)	Beschreibung
Situativ (interne/externe Variabilität)	Umgebung	(Wahrgenommener) Kontext der Interaktion
Vermittelt (initial/dynamisch)	Technisches System	Basiert auf vergangenen Erfahrungswerten (u. a. mit dem Automationssystem)

Die situativen Faktoren lassen sich wiederum unterteilen in interne Faktoren, die die beteiligte Person betreffen, und externe Faktoren, die die Begleitumstände betreffen. Als interne situationsbezogene Faktoren gelten Eigenschaften der Operatorin oder des Operators, die sich im Gegensatz zu den dispositionalen Faktoren erst aus der Situation heraus ergeben, z. B. das aktuelle Selbstvertrauen in Anbetracht der Arbeitsaufgabe oder die aktuelle Stimmung. Als externe Faktoren lassen sich bspw. die objektive Komplexität der Aufgabe oder die aktuelle Arbeitsbelastung bestimmen. Besonders interessant ist hierbei, dass diese situationsbezogenen Faktoren die Stärke des Zusammenhangs zwischen Vertrauen und Sich-Verlassen moderieren. Wenn bspw. Personen sich nicht kompetent darin fühlen, die relativen Vorteile durch die Nutzung einer Technologie einzuschätzen, kann dies dazu führen, dass diese dem System zwar vertrauen, aber nicht bereit sind, sich darauf auch zu verlassen. Gleiches gilt, wenn die Freiheitsgrade einer eigenständigen Entscheidung situativ eingeschränkt sind. Bei einer hohen mentalen Belastung kann bspw. das Sich-Verlassen auf die Technologie als einzige Möglichkeit zur Belastungsreduktion gesehen werden, sodass sich eine Person auf die Technologie verlässt, obwohl sie dieser nur wenig vertraut (Hoff & Bashir, 2015).

Abbildung 3.6 zeigt das komplette Drei-Schichten-Modell. Darin wird deutlich, dass vermittelte Faktoren zeitpunktabhängig in initiale Faktoren vor der Interaktion und dynamische Faktoren während der Interaktion unterteilt werden. Bereits vor der Interaktion entwickelt die beteiligte Person eine Erwartungshaltung hinsichtlich der Aktionen eines Roboters während der Interaktion. Als Determinante des initialen vermittelten Vertrauens nennen Hoff und Bashir (2015) das Vorwissen (*pre-existing knowledge*) dieser Person. Bestandteile dieses Vorwissens sind beispielsweise die vorhandenen Einstellungen und Erwartungen, die Reputation des Systems, Vorerfahrungen mit diesem oder ähnlichen Systemen sowie das Verständnis des Systems. Die Einstellungen können dabei auch implizit durch Assoziationen entstanden sein und unbewusst auf die Vertrauensbildung wirken. Insgesamt scheint es sich bei diesen Aspekten um kein (gesichertes) Wissen zu handeln, sondern um Erwartungen und Vorannahmen, die sich aus subtilen und unvollständigen Erfahrungswerten, aus mit Unsicherheit behafteten Analogieschlüssen und aus Aussagen Dritter

zusammensetzen (Kramer, 1999). Dies spiegelt sich auch in der ursprünglichen Definition nach Marsh und Dibben (2003, S. 469) wider, in der vermitteltes Vertrauen als „Erfahrung entstanden aus einer Sammlung vergangener situativer Vertrauenszustände“ (*experience born of a collection of past situational trusts*) definiert wird. Insofern wäre der Begriff des Vorwissens stark einzuschränken bzw. durch den treffenderen Begriff der *Vorerwartungen basierend auf mentalen Modellen* zu ersetzen, auf den die Autor:innen allerdings nicht zurückgreifen. Dass die Vertrauensbildung gerade in früheren Phasen auch durch analogische Prozesse beeinflusst ist, unterstreicht die Relevanz mentaler Modelle in diesem Zusammenhang. Das aktivierte mentale Modell bildet dabei u. a. die Zugehörigkeit zur Kategorie der menschlichen Wesen bzw. der Maschinen ab und evoziert damit ein spezifisches kategorienbasiertes Vertrauen (Kramer, 1999; J. D. Lee & See, 2004). Da das initiale Vertrauen allein durch Informationen Dritter vermittelt sein kann, lässt es sich synonym als *a priori*-Vertrauen bezeichnen, während das durch eigene Erfahrungen vermittelte Vertrauen während einer Interaktion als *a posteriori*-Vertrauen bezeichnet werden kann.

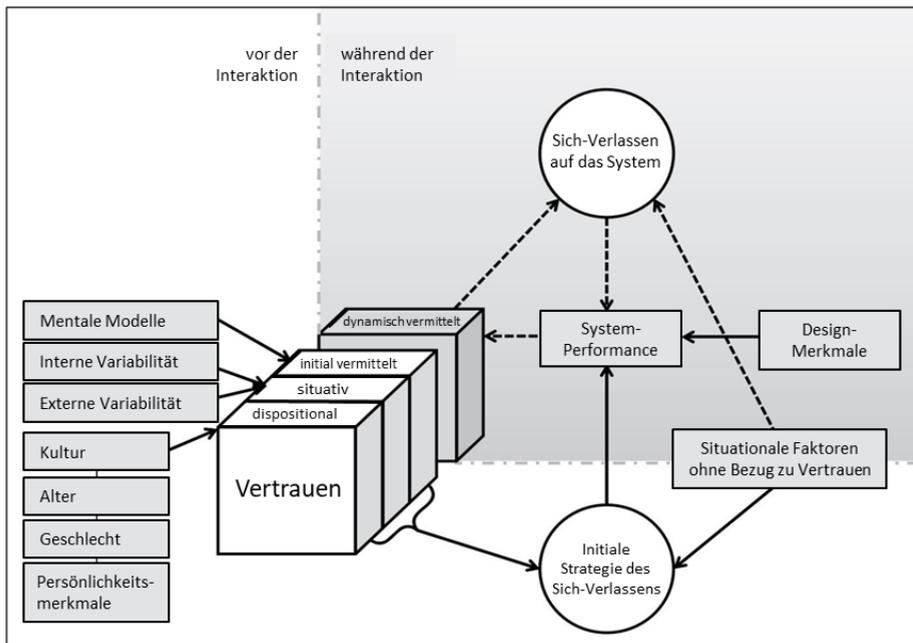


Abbildung 3.6: Schematische Darstellung des Drei-Schichten-Vertrauensmodells nach Hoff und Bashir (2015, S. 427). Eigene Übersetzung (Haftungsausschluss siehe Anhang A). Copyright © 2014 Human Factors and Ergonomics Society. Nachdruck mit Genehmigung von SAGE Publications.

Während der Interaktion mit dem System entstehen beständig neue Nutzungserfahrungen, die einen Abgleich der durch praktische Erfahrungen vermittelten Vertrauenswürdigkeit

mit dem aktuellen Vertrauensniveau ermöglichen. Das dynamische Vertrauen ist durch die (wahrgenommene) System-Performance bedingt, die von den Design-Merkmalen der Technologie abhängt. Qualitative Studien belegen, dass sich die relevanten Informationsquellen, Argumentationsmuster und Einflussfaktoren in Bezug auf initiales und dynamisches Vertrauen unterscheiden (Meißner et al., 2020). Das dynamische Vertrauen bestimmt das Maß, in dem menschliche Interaktionsteilnehmende dazu bereit sind, sich auf das System zu verlassen. Ob sie dies in angemessenem oder dysfunktionalem Maße tun, beeinflusst wiederum die tatsächliche Performance des Mensch-Roboter-Arbeitssystems. Damit ergibt sich ein zirkulärer Zusammenhang zwischen dem Vertrauen, dem Sich-Verlassen auf ein System und dessen Performance. Ferner wirken sich weitere situative Faktoren, wie z. B. der Aufwand, sich mit der Technologie zu beschäftigen, zeitliche Einschränkungen oder verfügbare Alternativen, auf den Grad des Sich-Verlassens auf die Technologie aus (Hoff & Bashir, 2015). Des Weiteren integrieren Hoff und Bashir (2015) Design-Eigenschaften des Cobots in ihr Modell, die sich auf die System-Performance, allerdings nicht direkt auf das Vertrauen auswirken. Damit ist das Modell nicht imstande, den möglichen direkten Einfluss eines anthropomorphen Roboter-Designs auf das empfundene Vertrauen abzubilden (Hoff & Bashir, 2015, S. 422).

Trotz dieser punktuellen Unklarheiten überzeugt das Modell insgesamt durch seine solide empirische Basis, die eine Vielzahl relevanter Einflussfaktoren einbezieht, durch seine Schlichtheit und die sowohl inhaltliche als auch zeitliche Differenzierung der Vertrauensfaktoren sowie durch die Darstellung der entsprechenden Rückkopplungsschleifen. Aufgrund der unterschiedlichen experimentellen Designs der in die Meta-Analyse einbezogenen Publikationen lässt sich die Stärke des Einflusses gewisser Elemente auf das resultierende Gesamtvertrauen nicht quantifizieren (Hoff & Bashir, 2015, S. 429).

3.4.7.3 Drei-Faktoren-Modell nach Hancock, Billings, Schaefer, et al. (2011)

Das sog. *Drei-Faktoren-Modell*¹³ nach Hancock, Billings und Schaefer, et al. (2011) und Hancock et al. (2020) modelliert im Gegensatz zu dem zuvor vorgestellten Ansatz nicht das Zusammenspiel einzelner vertrauensbildender Faktoren und deren zeitliche Dynamik, sondern beschreibt aus einer statischen Perspektive die zentralen Einflussfaktoren auf das Vertrauen in Roboter. Diese basiert auf einer viel beachteten und exklusiv auf Vertrauen in Roboter bezogenen quantitativen Meta-Analyse. Hancock, Billings und Schaefer, et al. (2011) unterteilen dabei die in bisherigen Studien untersuchten Einflussfaktoren auf Vertrauen in drei Gruppen, nämlich menschen-, roboter- und umweltbezogene Faktoren. Diese grundsätzliche Unterteilung entspricht der Differenzierung, die auch Hoff und Bashir (2015) ihrem Modell zugrunde legen. Insbesondere die Unterscheidung in roboter- und

¹³ Die Bezeichnung wurde u. a. in Billings, Schaefer, Chen und Hancock (2012, S. 110) und Schaefer, Chen, Szalma und Hancock (2016) verwendet, taucht allerdings nicht in den Primärquellen auf.

mitarbeiterbezogene Faktoren ist in vielen MRI-Studien gängig, wobei typischerweise u. a. die persönliche Vertrauensdisposition und die Grundeinstellung gegenüber Robotern als mitarbeiterbezogene Einflussfaktoren angeführt werden (You & Robert, 2018). Als zentrales Ergebnis der Meta-Analyse stellten Hancock, Billings und Schaefer, et al. (2011) fest, dass Vertrauen zum überwiegenden Teil durch die roboterbezogenen Faktoren bestimmt wird (Effektgröße $d = 0.47$). Ein Einfluss menschbezogener Faktoren ließ sich nicht nachweisen; die Effektgröße betrug $d = -0.02$ (Hancock, Billings, Schaefer et al., 2011, S. 523). Allerdings basierte diese Berechnung auf nur zwei Studien, die zum damaligen Zeitpunkt als die einzigen aussagekräftigen Forschungsarbeiten identifiziert wurden. Entsprechend begründeten die Autor:innen die Abwesenheit statistischer Effekte mit der unzureichenden Studienlage (Hancock, Billings, Schaefer et al., 2011, S. 523; Schaefer, Chen, Szalma & Hancock, 2016). Grundsätzlich wurden in der Folge die wenig eindeutige Zuordnung der Variablen zu den gebildeten Kategorien sowie die unzureichende Basis der Meta-Studie kritisiert, die aus nur 21 mitunter bis dato wenig beachteten Studien mit teils mangelhafter ökologischer Validität bestand (Cameron et al., 2015, S. 2). Die unterschiedliche Verfügbarkeit von Studien zu mensch- bzw. roboterbezogenen Faktoren zeigte sich auch in einer folgenden Meta-Analyse von Schaefer et al. (2016), die sich allgemeiner auf Literatur zu Vertrauen in Automation bezog. Diese Analyse bezog insgesamt neun Studien zu personenbezogenen Vertrauensfaktoren im Vergleich zu 22 Studien zu automationsbezogenen Faktoren ein. Die errechneten Effektstärken für die beiden Faktorengruppen unterschieden sich nur noch geringfügig (Hedges $g = 0.49$ bzw. 0.53). Nach Meinung der Autor:innen übt der Charakter der Nutzenden *a priori* den größten Einfluss auf das Vertrauen aus (Schaefer et al., 2016, S. 380).

In einer jüngeren Meta-Analyse (Hancock et al., 2020) zeigte sich je nach zugrundeliegendem statistischen Zusammenhangsmaß ein stärkerer Einfluss von personen- bzw. roboterbezogenen Einflussfaktoren. Die Autor:innen schlussfolgerten zwar, dass roboterbezogene Faktoren grundsätzlich einen größeren Einfluss als personenbezogene Faktoren haben. Diese Aussage spiegelt sich allerdings in den statistischen Kenngrößen nur bedingt wider. Ferner hoben die Autor:innen den Einfluss der Zuverlässigkeit des Roboters hervor, der während der Interaktion relevant ist und auch in den dynamischen Vertrauensmodellen berücksichtigt wird. Vorherige Erfahrungswerte mit Robotern ($d = 0.46$) und Anthropomorphismus ($d = 0.37$) zeigen ebenfalls einen kleinen bis mittelstarken positiven Effekt auf das Vertrauen. Zwar gelten gerade Vorerfahrungen allgemein als relevant für das Robotervertrauen, allerdings lässt sich nicht differenzieren, ob Vorerfahrungen zu einem höheren Vertrauen führen oder ob ein bereits vorhandenes Vertrauen den Erwerb von Erfahrungen ermöglicht. Außerdem wird in der Regel nicht einbezogen, ob diese Nutzungserfahrungen positiv oder negativ bewertet wurden (Sanders et al., 2017).

Tabelle 14: Vergleich der in unterschiedlichen Meta-Analysen ermittelten Effektstärken für den Einfluss bestimmter Faktorentypen auf Mensch-Roboter-Vertrauen.

	Hancock, Billings und Schaefer, et al. (2011)		Schaefer et al. (2016)	Hancock et al. (2020)	
	r	Cohens d	Hedges g	r	Cohens d
Personenbezogene Faktoren	+0.09	-0.02	+0.49	+0.12	+0.42
Umweltbezogene Faktoren	+0.11	+0.47	⁻¹⁴	+0.08	+0.54
Roboterbezogene Faktoren	+0.24	+0.67	+0.53 ¹⁵	+0.20	+0.32

Tabelle 14 stellt abschließend die Ergebnisse der unterschiedlichen Meta-Analysen gegenüber. Die starken Schwankungen je nach einbezogenen Studien und die zuvor thematisierten methodischen Schwächen der quantitativen Meta-Analyse erschweren eindeutige Interpretationen. Letztlich kann bis dato keine belastbare Aussage darüber getroffen werden, welche Gruppen an Einflussfaktoren die größte Auswirkung auf die Vertrauensbildung haben. Der oftmals vermittelte Eindruck, dass roboterbezogene Eigenschaften am einflussreichsten sind, könnte durch die höhere Verfügbarkeit von empirischen Ergebnissen begründet sein (Sanders et al., 2017, S. 1809).

3.4.7.4 Integriertes MRI-Vertrauensmodell

Hinweis: Einige Inhalte aus diesem Kapitel (inklusive der Abbildung) sind in Kopp, Baumgartner und Kinkel (2022) veröffentlicht.

Abbildung 3.7 stellt das im Rahmen dieser Dissertation entwickelte integrierte Vertrauensmodell für Mensch-Roboter-Vertrauen dar, das die relevanten Elemente der zuvor vorgestellten Modelle sowie weitere Vertrauensdifferenzierungen einbezieht. Das Kernstück bilden die verschiedenen Schichten des Vertrauens nach Hoff und Bashir (2015). Vor der Interaktion steht das kognitive Vertrauen im Vordergrund, wohingegen während der Interaktion affektive Vertrauenskomponenten eine maßgebliche, wenngleich nicht die alleinige Rolle spielen. Den unterschiedlichen Vertrauensschichten lassen sich die mensch- und die umweltbezogenen Faktoren aus dem Drei-Faktoren-Modell (Hancock, Billings, Schaefer et al., 2011; Hancock et al., 2020) zuordnen. So umschreiben die zeitstabilen menschbezogenen Faktoren wie grundsätzliche Charaktereigenschaften die individuelle Vertrauensdisposition. Situativ variierende menschbezogene Faktoren wie die aktuelle kognitive

¹⁴ Die Autor:innen konnten keine in diese Kategorie passenden Studien identifizieren.

¹⁵ Der angegebene Wert bezieht sich Automationstechnologie.

Belastung und Umweltfaktoren wie die Aufgabenkomplexität bedingen das situative Vertrauen. In gewissem Maße wirken auch bestehende mentale Modelle auf das situative Vertrauen, indem sie die Wahrnehmung der Interaktionssituation beeinflussen. Das vermittelte Vertrauen ergibt sich im Wesentlichen aus den vorhandenen mentalen Modellen, die die beteiligte vertrauende Person vom Roboter und ihrer Relation zu demselben hat. Der individuelle Anthropomorphismus sowie kursierende Ängste vor technologischer Ersetzung prägen dabei die vorherrschenden mentalen Modelle. Diese können durch sprachliches Framing beeinflusst werden und wirken im Rahmen von Top-Down-Prozessen auf die Wahrnehmung des Roboters und der MRI. Letztere wird zusätzlich von den während der MRI tatsächlich beobachtbaren Robotereigenschaften geprägt, die im Rahmen von Bottom-Up-Prozessen verarbeitet werden. Damit visualisiert das Vertrauensmodell explizit die beiden Ansatzpunkte für experimentelle Manipulationen. Dazu gehören einerseits das Framing, das auf hoher Verarbeitungsebene bei den mentalen Modellen ansetzt, und andererseits die Roboterkonfiguration, die die beobachtbaren Robotermerkmale und damit die Wahrnehmung des Roboters variiert.

Die Konzeption des Vertrauens als eine Abfolge von mentalen Zuständen basiert auf der TRA (Fishbein & Ajzen, 1975), gemäß derer aus der Vertrauenseinstellung die Bildung einer Handlungsabsicht (Mayer et al., 1995) und letztlich eines handlungsbezogenen Vertrauens (Hancock, Billings, Schaefer et al., 2011; Hancock et al., 2020) folgt. Das Maß, zu dem sich eine vertrauende Person auf eine Technologie verlässt, bzw. deren Ablehnung der Technologienutzung wird hierbei als handlungsbasierter Ausdruck von Vertrauen bzw. von Misstrauen verstanden. Die integrierten schematischen Diagramme verweisen jeweils darauf, dass neben einem optimalen Vertrauensniveau, das sich aus einer perfekten Übereinstimmung zwischen entgegengebrachtem Vertrauen und tatsächlicher Vertrauenswürdigkeit der vertrauensnehmenden Instanz ergibt, jeweils unvorteilhafte Über- und Unterschreitungen denkbar sind (vgl. Kapitel 3.4.6). Im Rahmen der aus den zugrundeliegenden Vertrauensmodellen bekannten Rückkopplungsschleife wirkt sich die Benutzung beständig auf die Wahrnehmung des Roboters aus. Während zu Beginn das Aussehen im Vordergrund steht, ist davon auszugehen, dass im Verlauf der Interaktion besonders die wahrgenommene Zuverlässigkeit und Leistungsfähigkeit in den Fokus rückt. Neben objektiven roboterbezogenen Faktoren, die die tatsächliche Zuverlässigkeit bedingen, beinhaltet die *wahrgenommene* Zuverlässigkeit stets subjektive Beurteilungen, die in Wechselwirkung mit den vorhandenen mentalen Modellen und daraus abgeleiteten Erwartungen stehen. Verletzungen von hohen Erwartungshaltungen oder Fehleinschätzungen der tatsächlichen Aktionen eines Roboters können bspw. zu einer sehr niedrigen wahrgenommenen Zuverlässigkeit und damit zu einem starken Einbruch des Vertrauens führen. Umgekehrt können mit steigender Nutzungszeit die Erfahrungen auch zu Adaptionen an den mentalen Modellen führen. Die Beobachtungen der Roboteraktionen und die eigenen Nutzungserfahrungen beeinflussen wiederum das dynamische vermittelte Vertrauen und damit auch das Vertrauen insgesamt.

Anhand dieses integrierten Modells lässt sich der experimentelle Fokus der vorliegenden Dissertation näher bestimmen. Dieser liegt auf den bis dato weniger untersuchten Top-Down-Einflüssen auf die Vertrauensbildung, die von den vorherrschenden mentalen Strukturen bei den menschlichen Interaktionspartner:innen anstatt von den beobachtbaren Merkmalen des beteiligten Roboters ausgehen. Präziser gefasst, steht die Auswirkung der vorherrschenden mentalen Modelle auf die Bildung des initialen vermittelten Vertrauens in Cobots im Vordergrund der Untersuchung. In diesen mentalen Modellen spiegeln sich wiederum der individuelle Anthropomorphismus und die Psychologie der technologischen Ersetzung wider, insofern eine Neigung zur Vermenschlichung menschenähnliche mentale Modelle aktiviert und die individuell wahrgenommene Bedrohung des eigenen Arbeitsplatzes durch Technologie bestimmt, wie die Mitarbeitenden ihre Relation zum Cobot in den mentalen Modellen repräsentieren. Als Methode zur Beeinflussung des Vertrauens steht damit das sprachliche Framing im Mittelpunkt der Dissertation, mittels dessen die Aktivierung bestimmter mentaler Modelle begünstigt werden kann. Daraus ergeben sich hypothetisch Auswirkungen auf das insgesamt empfundene Vertrauen und damit auf das Verhalten gegenüber dem Cobot sowie auf die Wahrnehmung desselben während der MRI.

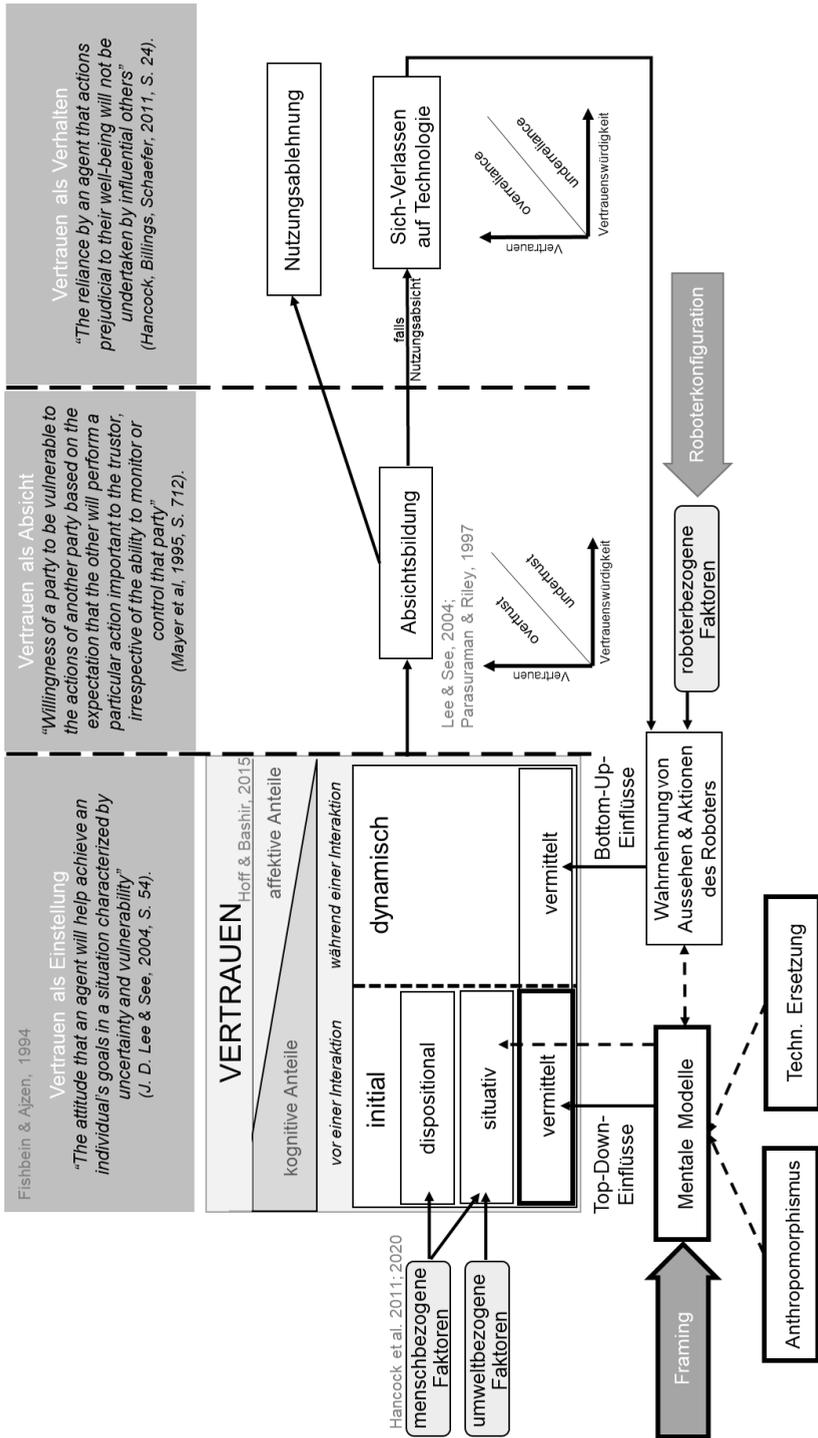


Abbildung 3.7: Integriertes Vertrauensmodell. Diejenigen Vertrauensaspekte, die in dieser Dissertation dezidiert untersucht werden, sind in der Abbildung fett umrandet.

3.4.8 Technologieakzeptanz

Während einige Forschungsarbeiten v. a. im Kontext von Automatisierungstechnologie Vertrauen als zentrales Kriterium für die MRI ins Zentrum rücken, operieren andere Forschungsarbeiten vorwiegend mit dem Konstrukt der Akzeptanz. Akzeptanz- und Vertrauensforschung sind mitunter nur sehr schwer voneinander abgrenzbar. Die Forschung zum Vertrauen in MRI wurde wesentlich durch das Modell von J. D. Lee und See (2004) geprägt, wohingegen die moderne Akzeptanzforschung ihre Wurzeln im Technologieakzeptanzmodell (TAM) nach Davis (1985) hat, das v. a. im Bereich der Akzeptanz von Informationssystemen von Bedeutung ist. Trotz ihres gemeinsamen Ursprungs in der TRA bzw. in der TPB koexistieren die Modelle zur Technologieakzeptanz bzw. zum Mensch-Technik-Vertrauen und werden von zwei Forschungsgemeinschaften je parallel weiterentwickelt (Ghazizadeh, Lee & Boyle, 2012). Mit der *unified theory of acceptance and use of technology* (UTAUT; Venkatesh, Morris, Davis & Davis, 2003) und der UTAUT2 (Venkatesh, Thong & Xu, 2012) bestehen allerdings auch Weiterentwicklungen des TAM, die den Anspruch erheben, über den Kontext der Informationstechnologie hinaus anwendbar zu sein.

Gemeinhin scheint Vertrauen das komplexere Konstrukt darzustellen, zumal Vertrauen in verschiedenen Disziplinen und Kontexten, wie z. B. dem gesellschaftlichen Zusammenleben, eine gewichtige Rolle spielt, wohingegen Akzeptanz schärfer umgrenzt ist. Ferner wird Akzeptanz häufig eher dichotom anstatt graduell interpretiert, d. h. es stellt sich eher die Frage, *ob* anstatt *wie sehr* eine bestimmte Technologie akzeptiert wird. Aufgrund derselben Basis in Form des TRA lässt sich auch Technologieakzeptanz als Einstellung, als Intention oder als Handlung begreifen (Müller-Abdelrazeq, 2020, S. 15). Die existierenden Technologieakzeptanzmodelle lassen jedoch zumeist eine präzise Bestimmung des Akzeptanzbegriffs vermissen und spezifizieren nicht, als welche Art von mentalem Zustand sie Akzeptanz verstehen. Oftmals wird implizit davon ausgegangen, dass sich die Akzeptanz in der tatsächlichen Handlung ausdrückt (vgl. Kritik in Meißner & Trübswetter, 2018, S. 224f.). Auch an dieser Stelle sei allerdings darauf verwiesen, dass im Bereich der industriellen MRI aufgrund der mangelnden Entscheidungsfreiheit der Mitarbeitenden u. U. eine erhebliche Differenz zwischen der tatsächlichen Nutzung eines Roboters und dessen Akzeptanz auf der Einstellungsebene bestehen kann (Oubari et al., 2018, S. 560f.). So kann es einerseits zu passiven Formen der Inakzeptanz auf der Einstellungsebene kommen, die sich nicht in den Handlungen der Person widerspiegeln, oder andererseits zu aktiven Formen der Inakzeptanz in Form eines Widerstands oder Boykotts der Technologie. Abbildung 3.8 zeigt die daraus resultierenden Formen der Akzeptanz. Befürwortung sowie (stille) Ablehnung bilden demnach die Gegensatzpaare für einstellungsbezogene und Engagement sowie Widerstand für handlungsbezogene Akzeptanz.

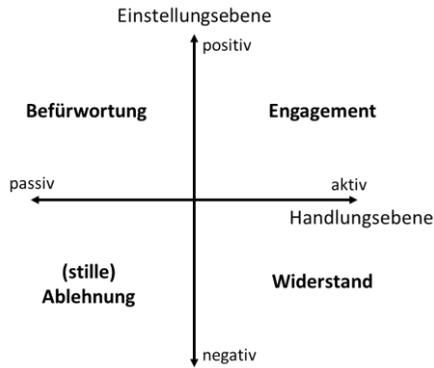


Abbildung 3.8: Formen der Akzeptanz auf Einstellungs- und Handlungsebene (Müller-Abdelrazeq, 2020, S. 16).

Der Ansatz der Akzeptanzmodelle besteht darin, Einflussgrößen auf die Technologienutzung zu bestimmen und diese innerhalb zumeist mehrstufiger Ursache-Wirkungs-Ketten zueinander in Beziehung zu setzen. Häufig werden anhand quantitativer Studien statistische Zusammenhangsmaße zwischen den einzelnen Faktoren errechnet. Der erste Entwurf des TAM (Davis, 1985) sah die Nutzungsabsicht durch die wahrgenommene Nützlichkeit und die wahrgenommene Benutzungsfreundlichkeit einer Technologie determiniert. Diese Ansicht blieb auch in den Weiterentwicklungen weitestgehend bestehen, wurde allerdings um weitere Einflussfaktoren ergänzt und insofern ausdifferenziert. Das war u. a. aufgrund des Einwands notwendig, dass sich TAM1 nicht auf einen verpflichtenden Nutzungskontext anwenden lässt, da sich in diesen Modellen die Nutzungsabsicht direkt aus der Einstellung ableitet, ohne dass externe Nutzungsverpflichtungen einbezogen werden (Ghazizadeh et al., 2012). Beim TAM2 (Venkatesh & Davis, 2000) wurden daher in erster Linie weitere Einflussfaktoren auf die wahrgenommene Nützlichkeit ergänzt, darunter die Freiwilligkeit der Nutzung. Im TAM3 (Venkatesh & Bala, 2008) finden sich zusätzlich Einflussfaktoren auf die wahrgenommene Benutzungsfreundlichkeit sowie weitere Faktoren, die in erster Linie den Zusammenhang zwischen diesen Konstrukten und der Nutzungsabsicht moderieren. Ferner wird eine Unterscheidung zwischen beeinflussbaren Justiervariablen (*adjustment variables*) und sog. Ankervariablen (*anchor variables*), also festen Einflussgrößen, eingeführt (Müller-Abdelrazeq, 2020, S. 28).

Das in Abbildung 3.9 dargestellte *automation acceptance model* (AAM) ist das einzige Modell, das das TAM explizit auf den Kontext der Automation überträgt und eine Beziehung zwischen Vertrauen und Akzeptanz modelliert (Ghazizadeh et al., 2012). Darin wirkt sich Vertrauen sowohl auf die wahrgenommene Benutzungsfreundlichkeit, die wahrgenommene Nützlichkeit als auch direkt auf die Nutzungsintention aus. Somit stellt Vertrauen einen Prädiktor für Akzeptanz dar, was zu der häufig vertretenen Auffassung passt,

dass Vertrauen eine Einstellung ist, wohingegen sich Akzeptanz in verschiedenen mentalen Zuständen ausdrücken kann (Müller-Abdelrazeq, 2020, S. 15). Obwohl sich Vertrauen in einigen anwendungsorientierten Studien tatsächlich als relevanter Prädiktor für die Akzeptanz erwies, beziehen andere Akzeptanzmodelle Vertrauen nicht explizit als Variable ein (Wu, Zhao, Zhu, Tan & Zheng, 2011).

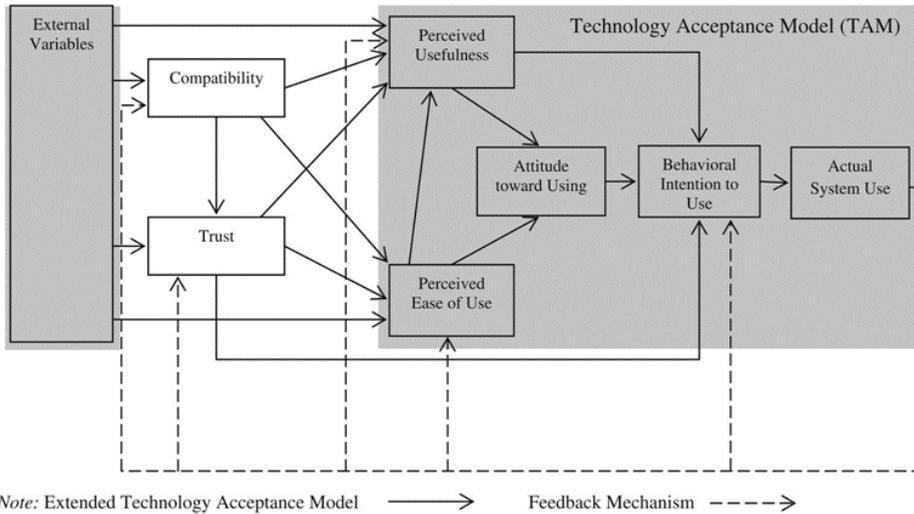


Abbildung 3.9: Vertrauen als Einflussvariable auf die Akzeptanz gemäß *automation acceptance model* (Ghazizadeh et al., 2012, S. 45). Copyright © Springer-Verlag London Limited 2011. Nachdruck mit Genehmigung von Springer Nature.

Inzwischen besteht mit dem TAM-HRC (*human-robot collaboration*) auch ein empirisch validiertes TAM, das dezidiert den MRI-Kontext adressiert (Bröhl, Nelles, Brandl, Mertens & Schlick, 2016, 2017; Bröhl, Nelles, Brandl, Mertens & Nitsch, 2019). Dieses Modell bietet eine detaillierte Übersicht über das Zusammenspiel und die Stärke des Zusammenhangs verschiedener Einflussfaktoren, wie in Abbildung 3.10 ersichtlich. Die darin enthaltenen Konstrukte wurden vorwiegend aus den klassischen TAM (Version 1-3) entnommen, auf den Kontext der Robotik adaptiert und punktuell um weitere Konstrukte ergänzt. Die ermittelten Korrelationskoeffizienten gelten als empirisch gut abgesichert, da sie mit großen Stichproben aus verschiedenen Kulturkreisen erhoben wurden. Da Akzeptanz innerhalb sozialer Kontexte konstruiert wird, sind kulturabhängig unterschiedliche Gewichtungen der Akzeptanzfaktoren naheliegend (Kamide & Arai, 2017, S. 537; Richert et al., 2017, S. 75; Vincent et al., 2015, S. 26), wengleich internationale Studien zeigen, dass geringfügige kulturelle Unterschiede, wie sie bspw. unter Bürger:innen der EU-

Mitgliedsstaaten vorhanden sind, im Vergleich zu individuellen Faktoren nur einen untergeordneten Einfluss ausüben (Turja & Oksanen, 2019).

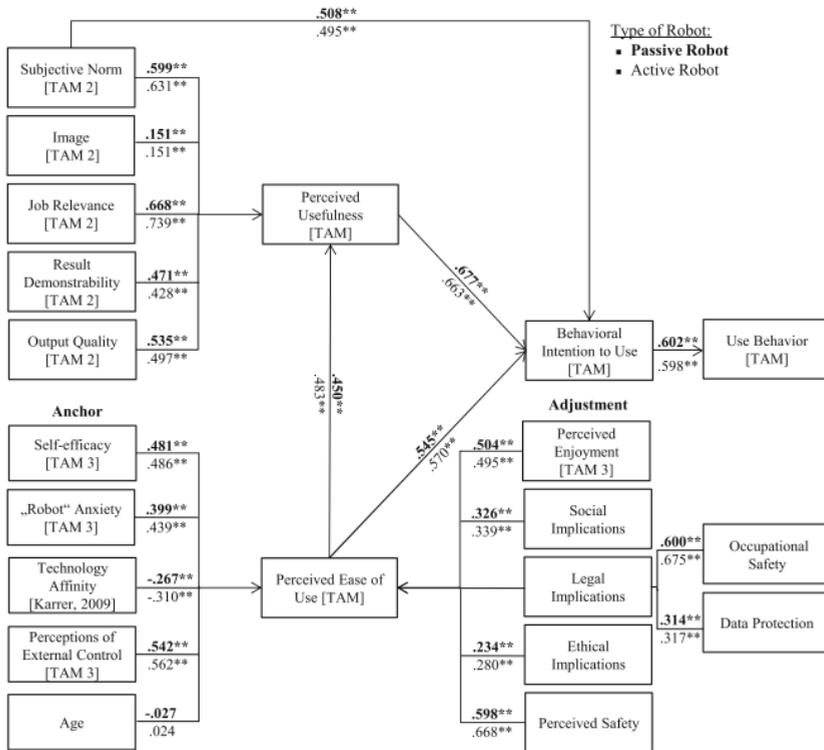


Abbildung 3.10: TAM-HRC für den MRI-Kontext mit empirisch ermittelten Zusammenhangsmaßen (Bröhl et al., 2019).

Die Erhebung der Einflussfaktoren über Single-Item-Abfragen stellt eine methodische Schwäche des TAM-HRC dar. Darüber hinaus problematisieren Meißner und Trübswetter (2018) die grundsätzliche Übertragbarkeit der TAM-Modelle auf den Roboterkontext, da sich Roboter zu stark von bestehenden Technologien unterscheiden und daher von Menschen völlig anders wahrgenommen werden. Außerdem sei in den Akzeptanzmodellen dem Umstand zu wenig Rechnung getragen, dass sich Akzeptanz zu einem großen Teil als Folge eines unbewussten, emotional beeinflussten und objektiv-rational nicht vollständig erklärbaren und kontextspezifisch unterschiedlich beeinflussten Prozesses ausbildet (Meißner & Trübswetter, 2018). Insofern eignen sich die dargestellten TAM und insbesondere das TAM-HRC zwar sehr gut zur Illustration einzelner Faktoren im Zusammenhang mit Technikakzeptanz und möglicher Operationalisierungen derselben, verkürzen aber gleichzeitig die Komplexität von Mensch-Technik-Relationen am Arbeitsplatz, indem sie

affektive Einflussfaktoren, subtile Wirkzusammenhänge und graduelle Ausprägungen weitestgehend außen vor lassen. Vor diesem Hintergrund bezieht sich die vorliegende Dissertation auf den umfassenderen Vertrauensbegriff, der trotz der vorhandenen Überschneidungen zur Akzeptanz besser für eine ganzheitliche Untersuchung der Mensch-Cobot-Relationen geeignet erscheint. Partiiell werden ergänzend einzelne Items aus den TAM bei der empirischen Erhebung einbezogen, wenn sie in dem jeweiligen Kontext besonders relevant erscheinen.

3.4.9 Vertrauensmessung

Ähnlich wie bei mentalen Modellen handelt es sich bei Vertrauen um ein latentes und damit nur schwer messbares Konstrukt (B. French et al., 2018, S. 49). Mensch-Technik- sowie Mensch-Roboter-Vertrauen werden selten verhaltensbasiert, sondern zumeist anhand subjektiver Selbsteinschätzungen in Multi-Item-Skalen gemessen, die allerdings das Risiko bergen, nur eine *ex post*-Rekonstruktion des Vertrauens abzubilden (Flook et al., 2019; B. French et al., 2018, S. 49; Spain et al., 2008). Häufig wurden die im MRI-Kontext verwendeten Skalen ursprünglich für die Messung von interpersonalem Vertrauen oder Mensch-Maschine-Vertrauen entwickelt (Schaefer, 2016). Ausnahmen bilden bspw. die *HRI trust scale* (Yagoda & Gillan, 2012) oder die *trust perception scale HRI* (Schaefer, 2016), die laut der Autorin insbesondere Unterschiede in Vorher-Nachher-Messungen vergleichsweise gut abbilden kann. Überdies findet die *trust in automation scale* (TAS; Jian, Bisantz & Drury, 2000) häufig Anwendung im MRI-Bereich (z. B. in Sanders et al., 2017) und gilt dort als eine der meistverwendeten empirisch validierten Multi-Item-Instrumente mit einer exzellenten internen Reliabilität (B. French et al., 2018; Hoff & Bashir, 2015). Während Kessler, Stowers, Brill & Hancock (2017) im MRI-Kontext die *HRI trust scale* anstatt der TAS empfehlen, schreiben Sanders et al. (2019) letzterer eine hohe Eignung im MRI-Bereich zu. Für die empirische Arbeit im Rahmen der vorliegenden Dissertation besteht ein zusätzlicher Vorteil der TAS darin, dass hierfür eine validierte deutschsprachige Version vorliegt (Pöhler, Heine & Deml, 2016). Daher wird in den folgenden Experimentalstudie die TAS zur Vertrauensmessung eingesetzt (vgl. Kapitel 4.5.2.3 und Kapitel 4.6.2.3).

Allen Messmethoden zum Vertrauen in Technologie gemein ist die Frage nach der Interpretierbarkeit absoluter Vertrauenswerte ohne konkreten Bezugspunkt. Neuere Studien operieren daher oftmals mit mehreren Messzeitpunkten und rücken zeitliche Vertrauensveränderungen stärker in den Fokus ihrer Untersuchungen (B. French et al., 2018, S. 49; Schaefer, 2016, S. 203). Doch selbst eine Vorher-Nachher-Messung ist kaum dazu in der Lage, sporadische und hochgradig dynamische Vertrauensveränderungen als Folge einzelner Ereignisse während einer Interaktion, wie z. B. unvorhergesehene Fehler eines Roboters, abzubilden (Desai et al., 2013). Hierfür wären kontinuierliche Vertrauensmessungen mit hoher zeitlicher Auflösung während der gesamten Interaktion notwendig, die mittels

integralbasierter Methoden wie AUC (*area under the curve*) zu einem Vertrauenswert verrechnet werden (Yang et al., 2017, S. 415). Das ist allerdings im Rahmen einer Vertrauensbestimmung auf Basis von Selbsteinschätzungen in der Praxis aufgrund zu erwartender erheblicher Ermüdungseffekte bei Mehrfachabfragen auf Seiten der Befragten kaum möglich. Doch auch die Verwendung verhaltensbasierter Messmethoden erweist sich als problematisch, weil sich aus beobachtbarem Verhalten kaum auf ein solch komplexes, soziales Phänomen wie Vertrauen rückschließen lässt, zumal weitere Einflussvariablen mit Vertrauen konfundieren (Flook et al., 2019). Eine ähnliche Herausforderung besteht bei neueren Messmethoden, die das zum jeweiligen Zeitpunkt bestehende Vertrauen aus neuronalen Gehirnaktivitäten abzuleiten versuchen. Bei den typischerweise kleinen Stichproben führen individuelle Einflussfaktoren auf Vertrauen dazu, dass es bisher nicht gelungen ist, eindeutige übergreifende neuronale Muster zu identifizieren (Nam, Eskander & Choo, 2021). Insgesamt lässt sich bisher ein Mangel an einheitlichen und konsistenten Messmethoden konstatieren, der die Vergleichbarkeit unterschiedlicher Studien einschränkt (B. French et al., 2018, S. 54).

3.4.10 Misstrauen

Die bisherigen Ausführungen zum Vertrauen sind ohne den Begriff des Misstrauens ausgekommen. Auch die meisten Forschungsarbeiten im MRI-Bereich fokussieren einzig auf den Vertrauensbegriff (B. French et al., 2018). Anhand des alltäglichen Sprachgebrauchs und aus dem Blickwinkel der Alltagspsychologie ließe sich Misstrauen als Gegenstück zu Vertrauen auffassen. In der Forschung wird das Verhältnis von Vertrauen und Misstrauen allerdings differenzierter betrachtet. Sowohl konzeptionelle als auch empirische Arbeiten liefern Hinweise darauf, dass es sich bei Vertrauen und Misstrauen nicht um die Endpunkte derselben Skala, sondern um zwei zwar miteinander in Verbindung stehende, aber grundsätzlich unterschiedliche Mechanismen und Konstrukte handelt (McKnight & Chervany, 2001; Sumpf, 2019, S. 86). Hierzu besteht innerhalb der Forschung derzeit jedoch noch kein Konsens (B. French et al., 2018).

Auch Luhmann betrachtet Vertrauen und Misstrauen als unterschiedliche Konstrukte. Diese These illustriert er anhand der Abwägung, einen Fernseher zu erwerben. Hierbei sei es zwar sinnvoll, eine Liste mit Vor- und Nachteilen der Anschaffung, nicht aber gleichzeitig eine solche Liste für den Fall der Nichtanschaffung anzulegen (Luhmann, 1968/2014, S. 92). Analog wäre es ausreichend, über wenig und viel Vertrauen zu sprechen und auf die Kategorie des Misstrauens komplett zu verzichten, sofern es sich beim Misstrauen nur um das Gegenstück zu Vertrauen handelte. Allerdings sei Misstrauen vielmehr ein funktionales Äquivalent des Vertrauens. Somit stehen Menschen zwei Alternativen zur Komplexitätsreduktion zur Verfügung, nämlich Vertrauen und Misstrauen. Wenn sich jemand dagegen entscheidet, Vertrauen zu schenken, stellt diese Person „die ursprüngliche

Komplexität der Geschehensmöglichkeiten wieder her und belastet sich damit“ (Luhmann, 1968/2014, S. 92f.). Um dieser belastenden Komplexität zu entgehen, muss die jeweilige Person auf eine andere funktional äquivalente Strategie zurückgreifen, nämlich das Misstrauen, das ebenfalls eine Festlegung auf eine bestimmte Zukunftserwartung impliziert. Nicht zu vertrauen, bedeutet gemäß der in Kapitel 3.4.3 hergeleiteten Vertrauensdefinition nicht zu erwarten, dass eine Technik bei der Erreichung eigener Ziele hilfreich sein wird. Es handelt sich also um eine Nicht-Erwartung eines positiven Zukunftsverlaufs. Zu misstrauen bedeutet hingegen, einen negativ bewerteten Zukunftsverlauf zu erwarten, d. h. zu erwarten, dass die Verwendung einer Technik der Erreichung eigener Ziele schadet.

In Bezug auf die Technikverwendung besteht die Komplexitätsreduktion beim Misstrauen im Verzicht auf die Nutzung oder zumindest auf eine Nutzung, die durch ein Sich-Verlassen auf die Technik gekennzeichnet ist.¹⁶ Eine Nutzung trotz hohen Misstrauens ist nur plausibel, falls diese durch externe Gründe gefordert wird. Bei dieser Beschreibung wird deutlich, dass Misstrauen (positive) Lernerfahrungen weitgehend ausschließt. Während mangelndes Vertrauen einen weiten Spielraum für mögliche Zukunftsverläufe lässt, bleibt im Falle des Misstrauens durch die damit verbundene Vermeidungsstrategie „wenig Raum (...) für unvoreingenommene, objektive Umwelterforschung und Anpassung“ (Luhmann, 1968/2014, S. 94). Daher sei es aus psychologischer Sicht vorteilhaft, eine Beziehung des Vertrauens gegenüber einer Beziehung des Misstrauens anzustreben. Das Vertrauen in eine Technologie führt je nach dessen Stärke und der damit zusammenhängenden Erwartung einer fehlerfreien Funktionsfähigkeit zu einem zunehmenden Verzicht auf die Kontrolle der Technologie während der Nutzung. Das setzt entsprechende kognitive Kapazitäten frei, die sonst aufgrund der permanenten Kontrolle gebunden wären, und kann zu einer gelingenden Interaktion beitragen. Über den erwähnten Nachteil der eingeschränkten Erfahrungsmöglichkeiten hinaus gibt es keinen Grund, Misstrauen gegenüber Vertrauen normativ abzuwerten oder als dysfunktional anzusehen, denn Misstrauen hat ebenso wie Vertrauen als komplexitätsreduzierender Mechanismus seine Daseinsberechtigung (Lewicki, McAllister & Bies, 1998). Abbildung 3.11 illustriert die Konsequenzen von Vertrauen und Misstrauen in Zusammenhang mit der empfundenen Komplexität der vorliegenden Situation.

In einer wegweisenden Arbeit schlossen Lewicki et al. (1998) an Luhmanns Argumentation an und postulierten eine zweidimensionale Auffassung von Vertrauen und Misstrauen. Sie argumentieren erstens damit, dass sich die beiden Konstrukte empirisch unterscheiden

¹⁶ Diese Aussage mag den Anschein erwecken, dass Misstrauen im Gegensatz zum graduellen Vertrauen eher als ein dichotomes Konzept verstanden wird, das unabhängig vom Grad des Misstrauens zur Konsequenz der Nichtbenutzung einer Technologie führt. Eine solch eindeutige Bestimmung ist allerdings fraglich und es sind durchaus Abstufungen im Grad des Misstrauens und der Verwendung einer Technologie denkbar, wenngleich die Unterscheidungen im Vergleich zum Vertrauen weniger feingranular erscheinen.

lassen und dabei unterschiedliche Entwicklungsmuster zeigen. Zweitens können Vertrauen und Misstrauen koexistieren und es entspreche der Erlebenswelt von Personen, anderen Personen in Hinblick auf unterschiedliche Vulnerabilitäten und Erwartungen gleichzeitig zu vertrauen und zu misstrauen. Dies erfordere eine separate Betrachtung der Antezedenten und Konsequenzen von Vertrauen und Misstrauen (Lewicki et al., 1998). Bezieht man dies auf die in Kapitel 3.4.3 eingeführte Definition nach J. D. Lee und See (2004), die Vertrauen als Erwartung bestimmt, dass ein:e Akteur:in bei der Erreichung individueller Ziele helfen wird, lässt sich diese Koexistenz leicht verdeutlichen. Da bspw. mit der Techniknutzung parallel mehrere womöglich sogar miteinander konfligierende Ziele verfolgt werden, kann die Technik für manche Ziele als hilfreich und für andere gleichzeitig als schädlich wahrgenommen werden. Das liefert gute Gründe für eine sowohl vertrauensvolle als auch misstrauische Haltung. Übereinstimmend gelangen McKnight und Chervany (2001) zu dem Schluss, dass die meisten Vertrauentheorien inzwischen Vertrauen und Misstrauen als unterschiedliche Konstrukte anstatt als die Endpunkte eines Kontinuums interpretieren.

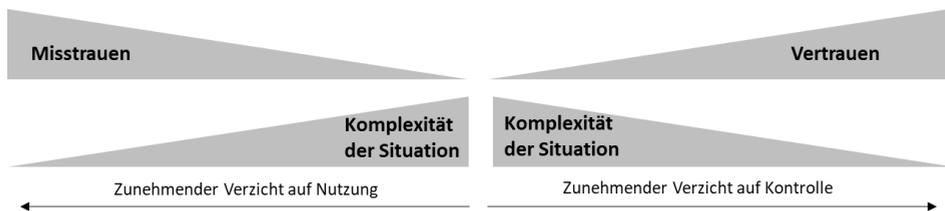


Abbildung 3.11: Vertrauen und Misstrauen als funktional äquivalente Mechanismen zur Reduktion von Komplexität.

Einige jüngere empirische Untersuchungen unterstützen diese Auffassung von Vertrauen und Misstrauen als zwei distinkte Konstrukte des menschlichen Erlebens. Hierbei wurde die populäre TAS (Jian et al., 2000) anhand konfirmatorischer Faktorenanalysen untersucht. Diese Skala enthält sowohl positiv formulierte Vertrauensitems als auch negierte Items, sodass sich bei der Auswertung einerseits durch Umkodierung der negativ formulierten Items ein eindimensionales Vertrauensmaß oder andererseits durch jeweils separate Auswertung der positiv und der negativ formulierten Items zwei getrennte Konstrukte bilden lassen. Während Jian et al. (2000) für eine eindimensionale Nutzung votierten, stellten Pöhler et al. (2016) und Spain et al. (2008, S. 4) fest, dass ein zweidimensionales Modell besser mit den empirischen Daten übereinstimmt (*model fit*). Weitere empirische Ergebnisse von Hergeth, Lorenz, Kreams & Toenert (2015) und McKnight und Chervany (2001, S. 27) stützen diese Ansicht. Im Bereich des Mensch-Technik-Vertrauens ist die Betrachtung von Misstrauen womöglich auch deshalb lohnenswerter, weil Menschen eher dazu neigen, ihr Misstrauen gegenüber Maschinen zum Ausdruck zu bringen, wohingegen sie

ihren Mitmenschen aufgrund sozialer Normen selten explizit das Misstrauen aussprechen (Madhavan & Wiegmann, 2007, S. 281).

Vertrauensverhältnisse können sich situationsbezogen in Misstrauensverhältnisse umkehren. Dabei löst nicht jede Irritation des Vertrauens direkt ein Umschlagen in ein Misstrauensverhältnis aus. Vielmehr lässt sich annehmen, dass Personen weitgehend indifferent auf solche Irritationen reagieren, sofern eine gewisse Schwelle nicht überschritten wurde, also redensartlich das Fass noch nicht zum Überlaufen gebracht worden ist: „Nicht jede Unstimmigkeit weckt Zweifel an den vertrauten Zügen der Umwelt, nicht jede Enttäuschung zerstört das Vertrauen. Eben deshalb muß es aber eine Grenze geben, wo diese Absorptionskraft endet, wo Vertrautheit oder Vertrauen abrupt in Mißtrauen umschlagen“ (Luhmann, 1968/2014, S. 96f.) Diese Indifferenz gegenüber Vertrauensverletzungen kann unter dem Begriff einer personenindividuellen Vertrauensresilienz gefasst werden, die im interpersonalen Vertrauen stärker als im Mensch-Maschine-Vertrauen ausgeprägt ist (de Visser et al., 2016). Wenn es allerdings zur Überschreitung der Toleranzschwelle für Vertrauensirritationen kommt, fällt die Reaktion oft heftig aus, da sich die vorher angestaute Spannung auf einen Schlag entlädt. Die Heftigkeit der Reaktion erscheint vor dem Hintergrund des konkreten Auslösers für die Schwellenüberschreitung oft unverhältnismäßig hoch. Überdies ist sie kaum vorhersagbar, da sich die Schwellen *ex ante* nicht exakt bestimmen lassen. Weiter erschwert wird eine solche Einschätzung dadurch, dass gewisse Ereignisse einen „symptomatischen Wert“ (Luhmann, 1968/2014, S. 99) zu gewinnen und damit das Vertrauens- oder Misstrauensverhältnis übermäßig stark zu verändern scheinen. In diesem Zusammenhang verweist Luhmann darauf, dass beinahe jede objektive Situation mögliche Gründe oder Beweise zur Rechtfertigung von Vertrauen oder Misstrauen bietet und oftmals eine „unbestimmte Vormeinung“ und der „Zufall eines ersten Eindrucks“ über die Tendenz bestimmt, welchem Komplexitätsreduktionsmechanismus man sich zuerst bedient (Luhmann, 1968/2014, S. 99). Erneut wird deutlich, dass Vertrauen und Misstrauen nicht oder zumindest nicht ausschließlich auf rationalen, sachlogischen Überlegungen beruhen.

Insgesamt lässt sich zusammenfassen, dass sowohl aus konzeptionell-soziologischer als auch aus empirisch-psychologischer Sicht deutliche Hinweise darauf bestehen, dass Vertrauen und Misstrauen stark voneinander abhängige, aber konzeptionell verschiedenartige Konzepte sind, die nicht als Endpunkte eines eindimensionalen Kontinuums verstanden werden sollten. Es bleibt allerdings fraglich, ob eine separate Betrachtung der stark miteinander korrelierenden Konstrukte im Rahmen anwendungsnaher Untersuchungen zu praxisrelevanten Unterschieden führt (vgl. z. B. Legler, Langer, Dittrich & Bullinger, 2020).

3.5 Zusammenführung und Forschungsprogramm

In den vorigen beiden Theoriekapiteln zur industriellen sowie zur kognitiven MRI wurden die zentralen theoretischen Grundlagen für die vorliegende Dissertation und die nachfolgenden empirischen Untersuchungen ausführlich dargelegt. Während die industrielle MRI den Anwendungskontext definiert, entstammen der kognitiven MRI mit der Theorie der mentalen Modelle, des Anthropomorphismus, des sprachlichen Framings sowie des Mensch-Roboter-Vertrauens die zentralen theoretischen Konzepte. Deren Zusammenspiel, das in Abbildung 3.12 visualisiert ist, wird in diesem Kapitel erläutert und zur Ableitung des zentralen Forschungsprogramms der vorliegenden Dissertation herangezogen.

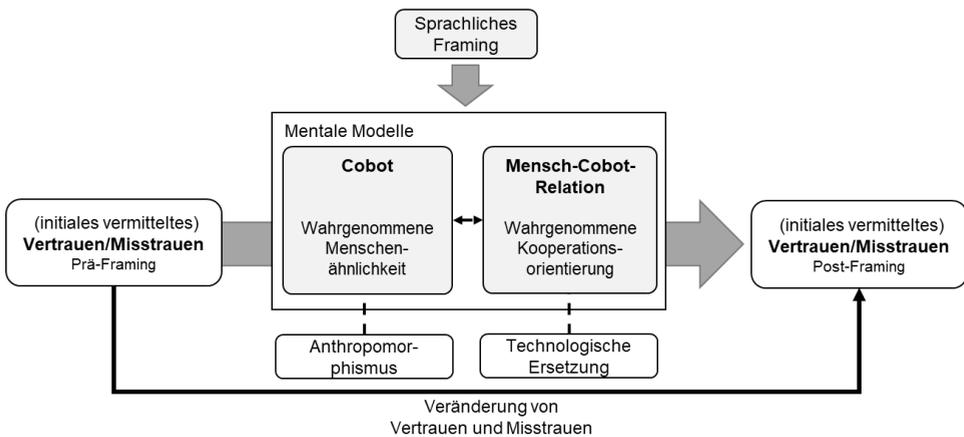


Abbildung 3.12: Zusammenspiel der theoretischen Konzepte und zentraler Forschungsansatz.

Wie die Ausführungen zur industriellen MRI zeigen, kann die Einführung hybrid gestalteter Cobots am industriellen Arbeitsplatz unter den dort beschäftigten Mitarbeitenden ambivalente Einstellungen hervorrufen. Diese basieren auf verschiedenen Möglichkeiten der mentalen Repräsentation des Cobots und der mentalen Konstruktion der Relation zu demselben. In Hinblick auf die mentalen Modelle von den Cobots stellt sich in erster Linie die Frage nach deren wahrgenommener Menschenähnlichkeit. Während der funktionale Charakter im Arbeitskontext für eine maschinenähnliche Repräsentation spricht, belegt die Forschung zum Anthropomorphismus, dass solche hybriden Roboter mitunter menschenähnlich repräsentiert werden, besonders insofern sie eine höhere Interaktivität erlauben und über anthropomorphe Gestaltungsmerkmale verfügen (vgl. Kapitel 2.2.4 und Kapitel 3.2). Ferner stellt sich vor dem Hintergrund der Angst vor Arbeitsplatzverlust die Frage nach der wahrgenommenen Kooperationsorientierung der konstruierten Mensch-Cobot-Relation. Eine kooperierende Repräsentation impliziert, dass der Roboter als willkommene

Unterstützung am Arbeitsplatz betrachtet wird, die sich positiv auf die Arbeitsergebnisse und -atmosphäre auswirkt. Eine konkurrenzorientierte Relation deutet hingegen an, dass der Roboter als bedrohliche:r Konkurrent:in mit überlegener Leistung erscheint. Aufgrund des schmalen Grats zwischen diesen beiden Repräsentationsformen und einer möglichen dynamischen Entwicklung von Unterstützung zu Ersetzung kann es zu häufigen Wechseln zwischen beiden Repräsentationen kommen. Das unterstützt die Annahme, dass sich die jeweilig aktivierten mentalen Modelle mittels sprachlichen Framings variieren lassen und sich in einem Top-Down-Prozess auf die Wahrnehmung einer MRI und das Verhalten innerhalb derselben auswirken (vgl. Kapitel 3.3.3). Das Forschungsvorhaben grenzt sich insofern von anderen Studien ab, die sich allein auf Bottom-Up-Prozesse fokussieren und dementsprechend wahrnehmbare Merkmale von Robotern wie z. B. deren Design variieren. Insbesondere ist auf Basis der Forschung zum Anthropomorphismus (vgl. Kapitel 3.2) und zur Psychologie der technologischen Ersetzung (vgl. Kapitel 2.3.2.2) anzunehmen, dass sich die vorherrschenden mentalen Modelle auf das empfundene Vertrauen auswirken, das im Arbeitskontext stets ein Vertrauen in den Erhalt des eigenen Arbeitsplatzes umfasst. Da sich die mentalen Repräsentationen besonders im Vorfeld einer MRI auswirken, solange noch keine realen Erfahrungswerte vorliegen, fokussiert die folgende empirische Arbeit die Einflüsse auf das initiale vermittelte Vertrauen.

Das Forschungsprogramm dieser Dissertation strukturiert sich entlang der folgenden sechs aufeinander aufbauenden Forschungsfragen (*FF*), die sich allesamt auf Cobots am industriellen Arbeitsplatz als Forschungsgegenstand beziehen.

FF1: Welche Rolle spielt Vertrauen als Erfolgsfaktor für betriebliche Cobot-Einführungen aus Sicht von Unternehmensvertreter:innen im Bereich der industriellen MRI?

Zwei anwendungsorientierte explorative Feldstudien sollen zunächst klären, welchen Stellenwert Unternehmensvertreter:innen im Bereich der industriellen MRI dem Vertrauen in Cobots beimessen, welche weiteren Erfolgsfaktoren sie für betriebliche Cobot-Einführungen als relevant ansehen und welche Rolle Vermenschlichung und Framing als mögliche Einflussfaktoren auf Vertrauen in diesem Kontext spielen. Die Studien werden unter Einsatz quantitativer und qualitativer Forschungsmethoden gemäß dem *mixed methods*-Ansatz durchgeführt und legen damit gleichzeitig die Grundlagen für die folgenden Experimentalstudien im Labor. Die Forschungsergebnisse dienen außerdem dazu, die aus Forschungssicht hergeleiteten theoretischen Grundlagen einer Praxissicht gegenüberzustellen, um eine ganzheitliche Perspektive zu erlangen. Die Relevanz der Fragestellung ergibt sich in erster Linie daraus, dass diese kognitiven Prozesse im industriellen Umfeld bisher nur wenig untersucht wurden.

FF2: Inwiefern lässt sich die Wahrnehmung eines Cobots und der Mensch-Cobot-Relation mittels sprachlichen Framings beeinflussen?

FF3: Wie wirken sich die wahrgenommene Menschenähnlichkeit eines Cobots und die wahrgenommene Kooperationsorientierung der Mensch-Cobot-Relation auf das Mensch-Cobot-Vertrauen aus?

Eine erste quantitative Experimentalstudie untersucht anschließend zwei aufeinander aufbauende Forschungsfragen. Zum einen steht dabei die Frage im Fokus (*FF2*), inwiefern sich mittels sprachlichen Framings die wahrgenommene Menschenähnlichkeit des Cobots und die wahrgenommene Kooperationsorientierung der Mensch-Cobot-Relation beeinflussen lassen. Die darauf aufbauende Forschungsfrage (*FF3*) beschäftigt sich damit, wie sich die durch das Framing veränderten mentalen Modelle auf das initiale vermittelte Vertrauen auswirken. Insofern trägt das Versuchsdesign dem bisher mangelnden Einbezug der Mensch-Roboter-Relation als mögliche Moderatorvariable auf die Einflüsse der wahrgenommenen Menschenähnlichkeit Rechnung.

FF4: Inwiefern wirken sich durch das Framing induzierte Veränderungen am initialen Vertrauen auf die reale Interaktion aus?

FF5: Inwiefern überdauern Framing-Effekte eine reale MRI?

Eine zweite quantitative Experimentalstudie beleuchtet weiterführende Forschungsfragen, die die Auswirkungen sowie zusätzliche Determinanten von Framing-Prozessen näher untersuchen. Um dem Bedarf nach realitätsnaher Forschung gerecht zu werden, wird in einem Präsenzexperiment untersucht, wie sich das Vertrauen auf eine reale MRI auswirkt (*FF4*) und inwiefern die Effekte des Framings eine reale MRI überdauern (*FF5*). Dabei steht das Zusammenspiel von kognitiven und affektiven sowie von initialen und dynamischen Vertrauenskomponenten im Fokus. Diese Verhältnisse sind sowohl aus Forschungsperspektive bspw. in Hinblick auf Modelle dynamischer Vertrauensentwicklung sowie aus praktischer Perspektive in Hinblick auf die nachhaltige Wirkung von Framing-Maßnahmen relevant, die bisher unzureichend untersucht ist (vgl. Kapitel 3.3.3).

FF6: Inwiefern reagieren verschiedene Populationen unterschiedlich auf das sprachliche Framing?

Da die beiden Experimentalstudien mit Stichproben aus unterschiedlichen Populationen durchgeführt werden, die mutmaßlich über eine unterschiedliche Einstellung zu Technik sowie über unterschiedliche Vorkenntnisse im Bereich der Robotik verfügen, lässt ein Vergleich der Ergebnisse Rückschlüsse darauf zu, welche Persönlichkeitsmerkmale die Wirksamkeit von Framing beeinflussen. Aus methodologischer Sicht erlaubt diese Gegenüberstellung Rückschlüsse auf die externe Validität vergleichbarer Studien, die zwar auf Studierende als Stichprobe zurückgreifen, die gewonnenen Erkenntnisse aber auf den Arbeitsplatzkontext verallgemeinern.

Abbildung 3.13 visualisiert das Forschungsprogramm und verdeutlicht die Abhängigkeiten zwischen den einzelnen Forschungsfragen und die Bezüge zum übergreifenden Forschungsdesign.

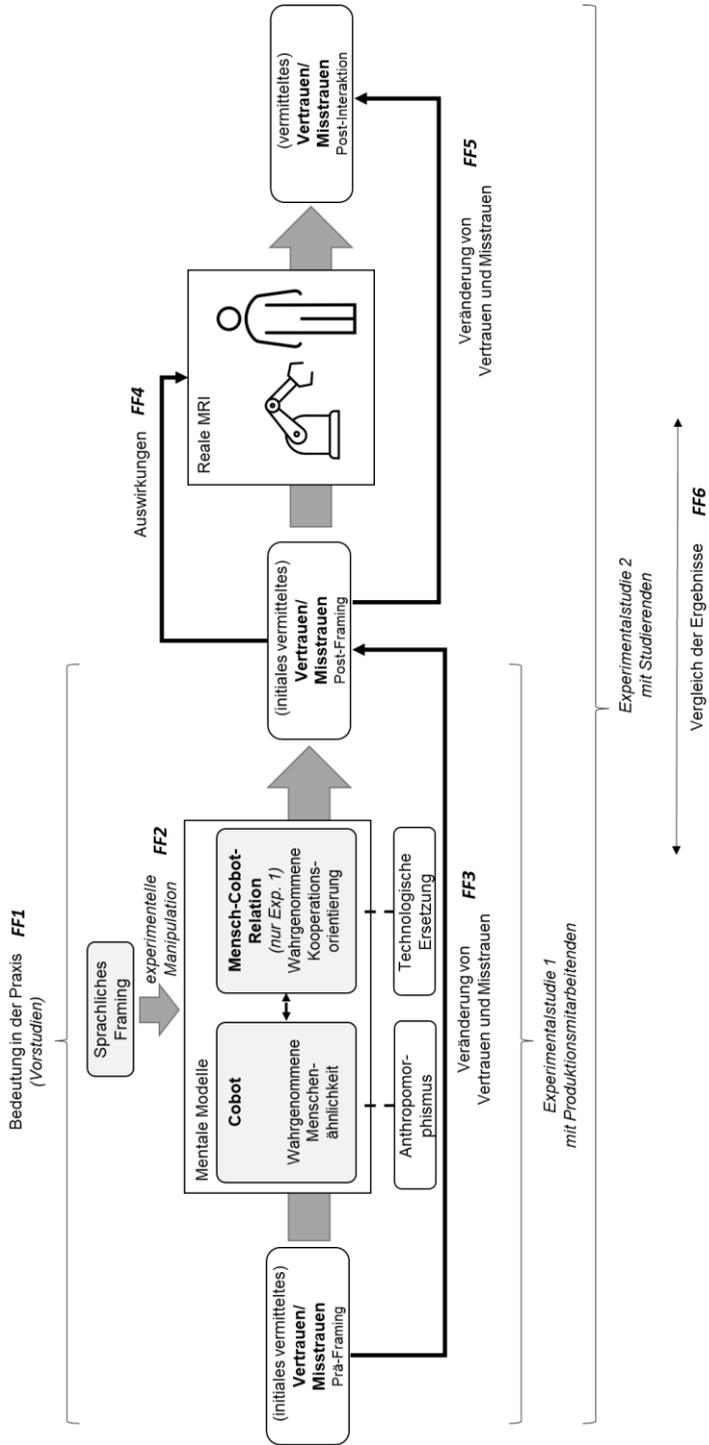


Abbildung 3.13: Übersicht über das Forschungsprogramm, strukturiert nach sechs zentralen Forschungsfragen, die in zwei Vorstudien und zwei Experimentalstudien untersucht werden.

4 Empirische Arbeit

In diesem Kapitel wird die empirische Arbeit im Rahmen der vorliegenden Dissertation dargestellt. Das einleitende Kapitel 4.1 schildert den grundsätzlichen methodischen Ansatz, der die vorliegenden Schwächen in bestehender MRI-Forschung zu kompensieren versucht. In Kapitel 4.2 und 4.3 werden zunächst zwei Vorstudien dargestellt, die sich der Beantwortung von *FF1* widmen. Danach folgt ein kurzes Zwischenfazit in Kapitel 4.4, bevor in den Kapiteln 4.5 und 4.6 die zwei darauf aufbauenden zentralen Experimentalstudien detailliert beschrieben werden, die sich mit den *FF2* bis *FF5* beschäftigen. Zur Klärung der übergeordneten *FF6* werden in Kapitel 4.7 die beiden Experimentalstudien verglichen. Eine kompakte Zusammenfassung der Ergebnisse aus der kompletten empirischen Arbeit entlang der Forschungsfragen findet sich in Kapitel 4.8.

4.1 Methodischer Ansatz

Der bestehende Mangel an umfassenden und empirisch abgesicherten Theorien im Bereich der kognitiven MRI lässt sich einerseits auf die Neuartigkeit der involvierten Technologien und andererseits auf die methodischen Einschränkungen bestehender Studien zurückführen. Die methodische Kritik bezieht sich u. a. auf die mangelnde Realitätsnähe, die sich in einer unzureichenden externen Validität der Studienergebnisse niederschlägt. So findet ein Großteil der Studien im Labor anstatt in der realen Anwendungsumgebung von Cobots, d. i. im Produktionsumfeld, statt und nutzt häufig übersimplifizierte Versuchsdesigns (Hegenberg et al., 2019; Jung & Hinds, 2018; Šabanović et al., 2014). Ferner werden zu meist Studierende, mitunter ausschließlich aus technischen Studiengängen, als Stichprobe herangezogen (*convenience sample*), was die Übertragbarkeit dieser Forschungsergebnisse auf die Zielpopulation von Produktionsmitarbeitenden in Zweifel zieht (Elprama et al., 2016; Horstmann et al., 2018; Müller-Abdelrazeq, Schönefeld, Haberstroh & Hees, 2019; Richert et al., 2017; Richert, 2018, S. 109). Hinzu kommt, dass die Stichprobengröße häufig keine statistisch signifikanten, repräsentativen Ergebnisse erlaubt (Bethel & Murphy, 2010; Savela et al., 2018; Wullenkord & Eyssel, 2019). Außerdem bemängeln Forschende die mangelnde Vergleichbarkeit bestehender Studien aufgrund uneinheitlicher Messprotokolle und einer unzureichenden Systematik der experimentellen Designs (Kompatsiari, Pérez-Osorio, Tommaso, Metta & Wykowska, 2018; Wiese et al., 2017). Bei den eingesetzten Forschungsmethoden dominieren bisher explizite Messverfahren zum Verhalten der Versuchsteilnehmenden wie z. B. Selbstauskünfte per Fragebogen, die aber im Vergleich zu impliziten Maßen zusätzlichen Verzerrungen (z. B. dem Effekt der sozialen Erwünschtheit) unterliegen und eine niedrigere „prädiktive[] Validität für das Verhalten“

(Wullenkord, 2017, S. 19) aufweisen (Bethel & Murphy, 2010; Stenzel, Chinellato, del Pobil et al., 2012; Wullenkord, 2017, S. 22). Entsprechend mehrten sich die Forderungen nach der Durchführung realitätsnaher *in situ*-Studien (Jung & Hinds, 2018; Onnasch & Roesler, 2019; Steil & Maier, 2017b, S. 417), die auf interdisziplinären Ansätzen beruhen und verschiedene (quantitative und qualitative) Forschungsmethoden kombinieren (Bethel & Murphy, 2010; Jung & Hinds, 2018; Moniz & Krings, 2016; Richert et al., 2017).

Aus methodischer Sicht verfolgt die vorliegende empirische Arbeit daher einen anwendungsnahen und methodenpluralistischen Ansatz, der insbesondere eine hohe externe Validität der Ergebnisse und eine ganzheitliche Untersuchung der komplexen Thematik anhand verschiedener Methoden anstrebt. Entsprechend verfolgen die nachfolgenden Vorstudien einen *mixed methods*-Ansatz, der Datenerhebungs- und Auswertungsmethoden aus der qualitativen und quantitativen Forschung miteinander kombiniert, um die jeweiligen spezifischen Schwachpunkte der Forschungsparadigmen möglichst zu kompensieren und auf Basis einer Triangulation der Ergebnisse ein ganzheitliches Bild des Forschungsgegenstands zu vermitteln (Kelle, 2019). Die beiden Vorstudien sowie die erste Experimentalstudie greifen auf Produktionsmitarbeitende als Untersuchungsgruppe zurück und die zweite Experimentalstudie integriert eine real stattfindende MRI, sodass möglichst realitätsnahe Gegebenheiten vorliegen und die Ergebnisse in hohem Maße auf den produktiven Cobot-Einsatz übertragbar sind. Ferner werden die relevanten psychologischen Konstrukte im Sinne einer hohen Validität und Reliabilität möglichst mittels anerkannter und validierter Selbstauskunftsskalen operationalisiert und ergänzend mittels impliziter Messmethoden erhoben.

4.2 Vorstudie 1: Online-Studie zu Erfolgsfaktoren für betriebliche Cobot-Einführungen

Hinweis: Einige Inhalte aus diesem Kapitel sind in Kopp, Baumgartner und Kinkel (2021) veröffentlicht.

4.2.1 Fragestellungen und Zielsetzung

Das Ziel der ersten Vorstudie bestand in der Identifikation von Erfolgsfaktoren für die Cobot-Einführung, um anschließend deren Relevanz aus der Sicht relevanter Unternehmensvertreter:innen zu eruieren. Ein besonderes Augenmerk lag dabei auf denjenigen Erfolgsfaktoren, die sich auf Vertrauen selbst, damit verwandte Aspekte oder potenziell vertrauensbildende Maßnahmen beziehen. Damit adressiert die erste Vorstudie folgende Forschungsfrage:

FFI: Welche Rolle spielt Vertrauen als Erfolgsfaktor für betriebliche Cobot-Einführungen aus Sicht von Unternehmensvertreter:innen im Bereich der industriellen MRI?

4.2.2 Identifikation und Strukturierung der Erfolgsfaktoren

Zur Identifikation möglicher Erfolgsfaktoren wurde im Oktober 2018 eine systematische Literaturanalyse durchgeführt. Aufgrund der Aktualität des Forschungsgebiets wurden hierbei nur Publikationen mit Veröffentlichungsdatum ab 2005 einbezogen. Die Suche bezog sich ausschließlich auf Artikel aus Fachzeitschriften, die das Peer-Review-Verfahren zur Qualitätssicherung anwenden. Die verwendeten Suchquellen und Suchterme lassen sich Tabelle 15 entnehmen. Ferner wurde die Literatursuche um das sog. Schneeballsystem bzw. die *Methode der konzentrischen Kreise* (Kornmeier, 2018, S. 89f.) erweitert, um auf Basis der Literaturverzeichnisse bereits identifizierter Publikationen weitere geeignete Veröffentlichungen zu identifizieren.

Tabelle 15: Verwendete Suchquellen und Suchterme bei der Literaturrecherche.

Suchquellen	Suchterm ¹
Literaturdatenbanken	
Springer Link	
Science Direct (Elsevier)	
Sage Journals	(collaborat* AND robot*) OR (cooperat* AND robot*) OR cobot
Taylor & Francis online	
DOAJ (Directory of open access journals)	
Meta-Suchmaschinen	
Google-Scholar	((collaborat* AND robot*) OR (cooperat* AND robot*) OR cobot) AND (manufacturing OR assembly OR industry OR industrial OR workplace)

Abbildung 4.1 veranschaulicht den gesamten Such- und Auswahlprozess. Die anfänglichen 754 Suchergebnisse wurden auf ihre Eignung geprüft und nach dem Schneeballsystem durch weitere 15 Publikationen ergänzt. Am Ende standen 37 geeignete wissenschaftliche Publikationen zur Verfügung, um darin enthaltene Erfolgsfaktoren für eine

¹ Der Asterisk in den Suchtermen fungiert als Platzhalter (*wildcard character*) für alle möglichen Zeichen.

betriebliche Cobot-Einführung zu identifizieren. Die Reduktion der Menge an Publikationen ergab sich durch das Entfernen von Dubletten sowie von Artikeln, die sich auf MRI-Anwendungen außerhalb des industriellen Bereichs fokussierten oder aufgrund ihres Fallstudiencharakters kaum verallgemeinerbare Informationen über relevante Erfolgsfaktoren beinhalteten.



Abbildung 4.1: Visualisierung der Prozessschritte bei der Literatursuche.

Um die Erfolgsfaktoren strukturieren zu können, wurde in einem nächsten konzeptionellen Schritt ein zweidimensionales Rahmenwerk entwickelt (vgl. Abbildung 4.2). Die vertikale Dimension des Rahmenwerks umfasst die einzelnen Bestandteile eines Mensch-Cobot-Arbeitssystems (Frennert, 2019). Angelehnt an das HTO-Framework (*human, technology, organisation*) nach Eklund (2003) ergeben sich als zentrale Betrachtungsgegenstände die interagierende Person (*human*), der Cobot (*technology*) und der organisationale Kontext (*organisation*), in dem die MRI stattfindet. Gerade die Relevanz der organisationalen und kontextuellen Faktoren wird von Trübswetter, Meißner, Weber, Klues und Stowasser (2018) im MRI-Kontext hervorgehoben. Diesen Komponenten werden die Erfolgsfaktoren nachfolgend zugeordnet. So wird bspw. der Erfolgsfaktor des Vertrauens in den Cobot der Ebene der interagierenden Personen zugeordnet, weil diese denjenigen Bestandteil des Arbeitssystems darstellen, der aktiv vertraut. Das Mensch-Cobot-System als solches wird als zusätzlicher vierter Betrachtungsgegenstand eingeführt, um diejenigen Faktoren zu fassen, die weder eindeutig dem Menschen noch dem Cobot, sondern nur dem System zuordenbar sind.

Auf der horizontalen Ebene wird gemäß dem typischen zeitlichen Verlauf einer Cobot-Einführung zwischen einer Entscheidungsphase, einer Planungs- und Implementierungsphase und einer Betriebsphase unterschieden. Während in der Entscheidungsphase prinzipiell über die mögliche Einführung eines Cobots entschieden wird, folgt in der Planungs- und Implementierungsphase bei positiver Entscheidung die genauere Spezifikation der Cobot-Installation und letztlich deren Implementierung. Während der Betriebsphase befindet sich der Cobot im alltäglichen, produktiven Einsatz und kann einer Leistungsevaluation unterzogen sowie bedarfsweise modifiziert werden. Hierin spiegeln sich auch die in Kapitel 2.3.2.1 besprochenen Phasen der Cobot-Einführung aus Sicht der Mitarbeitenden nach Wurhofer et al. (2015) wider, die um eine vorgelagerte Entscheidungsphase ergänzt wurden. Entscheidend für die Zuordnung der Erfolgsfaktoren zu den Phasen ist dabei das

Kriterium, in welcher Phase der jeweilige Faktor messbar wird. So lässt sich die tatsächliche Zuverlässigkeit des Cobots bspw. erst in der Betriebsphase empirisch bestimmen, wengleich dieser Faktor wie viele andere bestenfalls bereits in der Entscheidungsphase antizipiert wird. Das resultierende Rahmenwerk mit den 41 identifizierten und zugeordneten Faktoren ist in Abbildung 4.2 dargestellt und bildete die Basis für die folgende quantitative Online-Studie, die dazu diente, die wahrgenommene Relevanz der Faktoren in der unternehmerischen Praxis abzuleiten. Die unterschiedliche Anzahl an Faktoren in den einzelnen Zellen spiegelt nicht die Wichtigkeit der Betrachtungsdimensionen wider, sondern liefert ein Abbild davon, welche Faktortypen in der bisherigen Forschung thematisiert und untersucht wurden.

		Horizontale Dimension: Zeitverlauf		
		Entscheidungsphase	Planungs- & Implementierungsphase	Betriebsphase
Vertikale Dimension: Mensch-Cobot-Arbeitssystem	Mensch 	<ul style="list-style-type: none"> • Initiales Vertrauen (vor der Interaktion) • Hohes Selbstvertrauen • Vorwissen zu Industrierobotern • (Geeignete) Körpermaße der Mitarbeitenden • Vorwissen zu Roboterprogrammierung • Demografische Merkmale (Alter, Geschlecht, Kultur) 	<ul style="list-style-type: none"> • Eindruck, gut informiert zu sein • Keine Angst vor Arbeitsplatzverlust • Befürwortung der Einführung 	<ul style="list-style-type: none"> • Vertrauen während des Betriebs • Reduktion der mentalen Beanspruchung • Wahrnehmung des Cobots als Kollegen
	Cobot 		<ul style="list-style-type: none"> • Greifersystem • IT-Sicherheit • Intuitive Erkennbarkeit des Cobot-Status • Geschwindigkeit an Mitarbeitende angepasst • Geschwindigkeit möglichst hoch • Lärmvermeidung • Akustische Signale • Größe (Einschüchterung) • Erkennung menschlicher Zustände • Gefällige Gestaltung (Form, Farbe, Oberfläche etc.) • Vergabe eines menschl. Namens • Design (stark) menschenähnli. • Design (moderat) menschenähnli. 	<ul style="list-style-type: none"> • Zuverlässigkeit des Cobots • Vorhersagbarkeit der Cobot-Bewegungen
	Arbeits-system 	<ul style="list-style-type: none"> • Eignung des Produktionsprozesses • Mobilität und Umrüstbarkeit 	<ul style="list-style-type: none"> • Arbeitssicherheit • Aufgabenteilung • Standardisierung (Arbeitsabläufe/ Ansprechpartner) • Positionierung von Cobot und Arbeitsmaterial • Ergonomie 	<ul style="list-style-type: none"> • Standards bei Störungen
	Unter-nehmen 	<ul style="list-style-type: none"> • (Einmalige) Anschaffungskosten 	<ul style="list-style-type: none"> • Involvement des Managements • Kosten der Instandsetzung • Dauer der Instandsetzung • Einbezug der Gewerkschaften 	<ul style="list-style-type: none"> • (Fortlaufende) Betriebskosten

Abbildung 4.2: Entwickeltes Rahmenwerk zur Strukturierung der Erfolgsfaktoren für eine Cobot-Einführung.

4.2.3 Methoden und Material

4.2.3.1 Stichprobenbeschreibung

Die Teilnehmenden der Online-Studie wurden über den Panel-Anbieter *ToLuna Deutschland GmbH* aus der deutschen Bevölkerung rekrutiert. Um an der Befragung teilnehmen zu können, mussten die Teilnehmenden in einem Unternehmen aus der produzierenden Industrie in einer planenden oder leitenden Position (z. B. als Produktionsplaner:in, Manager:in, Teamleitung oder Betriebsleitung) im Bereich Fertigung/Produktion oder im Bereich Logistik/Materialien angestellt sein. Damit sollte im Sinne einer möglichst hohen externen Validität der Bewertungen sichergestellt werden, dass alle Befragten über ausreichendes praktisches Fachwissen in Bezug auf Produktionsprozesse verfügen und mit strategischen Überlegungen in diesem Kontext vertraut sind.

Im Rahmen der Datenbereinigung wurden sog. Durchklicker (*speeder*) zur Qualitätssteigerung aus dem Datensatz entfernt, wenn sie die komplette Studie in insgesamt weniger als 5 Minuten abschlossen. Diese geringe Bearbeitungsdauer deutet darauf hin, dass die Fragen ohne die notwendige Sorgfalt und willkürlich beantwortet wurden (Prill, 2017, S. 229). Der Median der Bearbeitungszeit der bereinigten Datensätze betrug $Mdn = 10:06$ Minuten ($M = 12:10$ Minuten, $SD = 7:58$ Minuten). Alle Teilnehmenden nahmen auf freiwilliger Basis teil und erhielten für das Ausfüllen der Umfrage gemäß den Nutzungsbedingungen des Panel-Anbieters eine Entlohnung.

Daraus ergab sich ein Stichprobenumfang von $n = 81$ Teilnehmenden (63 männlich, 18 weiblich). Diese waren zwischen 24 und 75 Jahre alt ($M = 46.96$, $SD = 12.75$) und verfügten über eine praktische Berufserfahrung zwischen einem und 47 Jahren ($M = 15.94$, $SD = 10.63$). 49 Teilnehmende hatten eine Lehre abgeschlossen (60.5 %), 30 verfügten über einen akademischen Hochschulabschluss (37.0 %) und zwei weitere hatten weder eine Lehre noch ein Studium abgeschlossen (2.5 %). 61 von 81 Teilnehmende (75.3 %) arbeiteten in der Produktion/Fertigung, während 14.8 % im Management und 9.9 % in der Logistik/Materialwirtschaft tätig waren. 85.3 % der Teilnehmenden hatten eine leitende Position inne, z. B. als Teamleitung oder Fabrikleitung, während die restlichen 14.8 % als Produktionsplaner:in arbeiteten. 38 von 81 (46.9 %) Befragten hatten sich bereits sehr intensiv mit Automatisierungslösungen beschäftigt, 16 von 81 (19.8 %) sogar dezidiert mit Cobots. 21 Teilnehmende (25.9 %) verfügten über praktische Erfahrungen mit der Einführung eines Cobots. Auf Grundlage dieser Daten lässt sich darauf schließen, dass die Teilnehmenden über ausreichende Erfahrung und Fachkenntnisse verfügten, um qualifizierte Informationen in Hinblick auf die relevanten Fragen zu liefern. 46 von 81 Teilnehmenden (57 %) arbeiteten in großen Unternehmen, während die restlichen 35 Teilnehmenden

(43 %) in KMU² beschäftigt waren. Große Unternehmen sind daher in der vorliegenden Studie leicht überrepräsentiert. Die beobachtete Branchenverteilung war insgesamt weitestgehend repräsentativ für die deutsche Wirtschaft (vgl. Anhang B.I.b).

4.2.3.2 Experimenteller Ablauf

Zu Beginn wurde mithilfe von drei Screening-Fragen geprüft, ob die Besucher:innen der Online-Studie die Kriterien für die Teilnahme erfüllten. Danach erhielten alle Teilnehmenden einige einführende Informationen über Cobots und wurden ohne deren Wissen quasi-zufällig einer Versuchsgruppe zugeordnet. Entsprechend dieser Zuordnung wurde den Teilnehmenden nachfolgend eine Beschreibung eines bestimmten fiktiven Unternehmensszenarios präsentiert, in das diese sich gedanklich hineinversetzen sollten (vgl. Kapitel 4.2.3.3). Um sicherzustellen, dass die Teilnehmenden die gesamte Beschreibung des Szenarios lasen, war es erst nach einer Verweilzeit von mindestens 20 Sekunden möglich, auf die nächste Seite zu gelangen.

Danach erfolgte eine Manipulationskontrolle (*manipulation check*). Die Teilnehmenden sollten hierzu das zuvor erwähnte strategische Ziel des fiktiven Unternehmens auswählen. Falsche Antworten führten zum Ausschluss von der Studie. Auf der nächsten Seite wurden die Teilnehmenden um eine Einschätzung der Realitätsnähe des präsentierten Szenarios gebeten. Danach sollten sie sich vorstellen, für einen Cobot-Einführungsprozess innerhalb des beschriebenen Unternehmens verantwortlich zu sein und in Freitextfeldern die aus ihrer Sicht wichtigsten Erfolgsfaktoren nennen. Daraufhin folgte die Bewertung der 41 Erfolgsfaktoren aus der Literaturanalyse, die in Blöcken auf vier Seiten aufgeteilt und in jeweils randomisierter Reihenfolge dargeboten wurden. Die Faktoren sollten jeweils nach ihrer Relevanz in Hinblick auf die Erreichung des Unternehmensziels aus dem fiktiven MRI-Szenario bewertet werden.

Am Ende der Studie erfolgte die Abfrage von deskriptiven Variablen zur Stichprobenbeschreibung. Nach dem Absenden ihrer Antworten gelangten die Versuchspersonen auf die Website des Panel-Anbieters, wo sie ihre Belohnung für die Teilnahme erhielten.

4.2.3.3 Materialien

Als Stimuli für die Studie dienten drei unterschiedliche textuell beschriebene Szenarien, die sich allesamt auf ein fiktives Unternehmen namens *Meyer AG* bezogen. Jedes Szenario umschrieb eine aktuelle Herausforderung, mit der sich das Unternehmen konfrontiert sieht, sowie ein sich daraus ergebendes spezifisches Ziel, das das Unternehmen durch die Cobot-Einführungen zu erreichen versucht (vgl. Tabelle 16). Die beschriebenen Zielsetzungen in

² Die Klassifikation in Groß-, mittlere und Kleinunternehmen erfolgte gemäß den Richtlinien der Europäischen Union (Europäische Union, 2003). Bei der Klassifizierung wurde die Bilanzsumme als Kriterium vernachlässigt und aus forschungspragmatischen Gründen nicht erhoben.

den drei Szenarien entsprechen den typischerweise in der Realität verfolgten Zielen von Unternehmen. Die beschriebenen Herausforderungen und Zielsetzungen in den verwendeten Szenarien sind in Tabelle 16 aufgelistet. In Anhang B.I.a ist überdies die Motivation zur Auswahl dieser Zielsetzungen ausführlicher dargelegt.

Tabelle 16: Beschriebene unternehmerische Herausforderungen und Zielsetzungen in den verwendeten Szenarien.

Herausforderung	Zielsetzung
Fachkräftemangel	Steigerung der Arbeitsplatzattraktivität
Erhöhter Bedarf an kundenindividueller Fertigung	Erhöhung der Flexibilität in der Fertigung
Hoher Kostendruck/Steigende Nachfrage	Erhöhung der Produktivität

4.2.3.4 Messmethodik

Die Relevanz der vorgegebenen Erfolgsfaktoren wurde anhand einer einseitigen fünfstufigen Skala beurteilt. Die verbalen Beschriftungen reichten gemäß der Empfehlung nach Rohrmann (1978) von *1 = nicht wichtig* bis *5 = sehr wichtig*. Aufgrund der äquidistanten verbalen Bezeichnungen wird die Skala für die Datenanalyse als quasi-metrisch behandelt.

Zur Einschätzung der Realitätsnähe der präsentierten Szenarien wurde ebenfalls eine einseitige fünfstufige Skala mit den verbalen Beschriftungen von *1 = nicht realistisch* bis *5 = sehr realistisch* eingesetzt. Anhand derselben sollten die Teilnehmenden bewerten, inwiefern sie es für realistisch hielten, dass sich ihr Arbeitgeber erstens ähnlichen Herausforderungen gegenüber sieht wie das fiktive Unternehmen und zweitens ebenfalls die Einführung einer Cobot-Lösung in Betracht ziehen würde, um den vorgenannten Herausforderungen zu begegnen.

4.2.3.5 Studiendesign

Die Studie folgt einem *between subject*-Design mit dem Zwischensubjektfaktor des verfolgten Ziels der Cobot-Einführung. Aufgrund unterschiedlich hoher Abbruchraten variierte die Anzahl vollständiger Datensätze pro Bedingung leicht. In der Bedingung *steigende Arbeitsattraktivität* befanden sich 28 Teilnehmende, in der Bedingung *steigende Flexibilität* 25 Teilnehmende und in der Bedingung *steigende Produktivität* weitere 28 Teilnehmende. Um den Einfluss des zugewiesenen Szenarios auf die Bewertung der Erfolgsfaktoren zu untersuchen, wurde eine multivariate Varianzanalyse (*multivariate analysis of variance*; MANOVA) durchgeführt, bei der die Versuchsbedingung als dreistufige nominale unabhängige Variable (UV) und die durchschnittliche Bewertung jedes Erfolgsfaktors als quasi-metrische abhängige Variable (AV) fungierte.

4.2.4 Ergebnisse und Diskussion

Die Versuchsteilnehmenden erachteten es als mehr als *einigermaßen realistisch*, dass sich das Unternehmen, bei dem sie arbeiten, mit der im jeweiligen Szenario beschriebenen Herausforderung konfrontiert sieht ($M = 3.20$, $SD = 1.24$) und dieser mit der Einführung eines Cobots begegnen würde ($M = 3.26$, $SD = 1.29$). Hinsichtlich dieser beiden Einschätzungen ergab eine einseitige Varianzanalyse (*analysis of variance*; ANOVA) keine signifikanten Unterschiede zwischen den drei unterschiedlichen Unternehmensszenarien, $F(78) = 2.247$, $p = .112$ und $F(78) = 1.771$, $p = .177$. Auf Basis dieser Ergebnisse lässt sich annehmen, dass die Teilnehmenden die Szenarien als ausreichend realitätsnah erachteten, um die Wichtigkeit der möglichen Erfolgsfaktoren profund beurteilen zu können.

In einer weiteren Analyse wurde anhand einer MANOVA geprüft, inwiefern sich die Bewertungen der Faktoren in Abhängigkeit vom präsentierten Unternehmensszenario unterschieden. Die MANOVA wurde angewendet, obwohl keine Normalverteilung der Daten gegeben war, da dieses Verfahren als robust gegenüber Verletzungen der Annahme einer Normalverteilung angesehen wird (Finch, 2005). Die Korrelationen zwischen den abhängigen Variablen waren ausreichend gering ($r < .90$), sodass keine Multikollinearität vorlag. Es wurden keine multivariaten Ausreißer auf Basis des Mahalanobis-Abstands³ identifiziert, $p > .001$. Die Analyse ergab keinen signifikanten Effekt, $F(82,76) = 1.05$, $p = .415$; Wilks' $\Lambda = .220$. Demnach ist die Relevanz der Erfolgsfaktoren insgesamt unabhängig von dem Ziel, das mit der Cobot-Einführung verfolgt wird. Entsprechend werden im Folgenden die jeweiligen Mittelwerte pro Erfolgsfaktor über alle drei Versuchsbedingungen hinweg zur Interpretation verwendet.

Abbildung 4.3 zeigt die durchschnittliche Relevanz, die die Befragten den ermittelten Faktoren für den Erfolg einer Cobot-Einführung beimaßen. Diejenigen Faktoren, die sich auf Vertrauen (gestrichelt-gepunktete Umrandung) sowie die hypothetisch damit in Verbindung stehenden Themenfelder des Anthropomorphismus (gepunktete Umrandung) und der Angst vor Arbeitsplatzverlust (gestrichelte Umrandung) beziehen, sind innerhalb des Rahmenwerks hervorgehoben. Die drei Faktoren mit der höchsten Relevanz pro Phase sind grau hinterlegt und gemäß ihrer Platzierung innerhalb der jeweiligen Phase nummeriert.

³ Multivariate Ausreißer weisen ein unübliches Muster an Werten für die verschiedenen AVs auf. Daher sind spezielle Methoden zur Identifikation notwendig, wie bspw. der Mahalanobis-Abstand, der den Abstand eines Datensatzes zum Zentroiden aller übrigen Datensätze unter Einbezug sämtlicher AVs berechnet. Vgl. hierzu Tinsley und Brown (2000, S. 13).

	Entscheidungsphase		Implementierungsphase		Betriebsphase	
Mensch	3. Initiales Vertrauen (vor der Interaktion)	4.05	3. Eindruck, gut informiert zu sein	4.75	3. Vertrauen während des Betriebs	4.48
	Hohes Selbstvertrauen	3.85	Keine Angst vor Arbeitsplatzverlust	4.59	Reduktion der mentalen Beanspruchung	3.25
	Vorwissen zu Industrierobotern	3.74	Befürwortung der Einführung	4.16	Wahrnehmung des Cobots als Kollegen	2.91
	(Geeignete) Körpermaße der Mitarbeitenden	3.57				
	Vorwissen zu Roboterprogrammierung	3.54				
	Demografische Merkmale (Alter, Geschlecht, Kultur)	3.23				
Cobot			2. Greifersystem	4.80	1. Zuverlässigkeit des Cobots	4.81
			IT-Sicherheit	4.62	Vorhersagbarkeit der Cobot-Bewegungen	4.43
			Intuitive Erkennbarkeit des Cobot-Status	4.16		
			Geschwindigkeit an Mitarbeitende angepasst	4.15		
			Geschwindigkeit möglichst hoch	4.02		
			Lärmvermeidung	4.00		
			Akustische Signale	3.89		
			Größe (Einschüchterung)	3.79		
			Erkennung menschlicher Zustände	2.75		
			Gefällige Gestaltung (Form, Farbe, Oberfläche etc.)	2.73		
			Bezeichnung mit menschlichem Namen	2.31		
			(stark) menschenähn. Cobot-Design	2.26		
			(moderat) menschenähn. Cobot-Design	2.15		
Arbeitssystem	1. Eignung des Produktionsprozesses	4.67	1. Arbeitssicherheit	4.83	2. Standards bei Störungen	4.54
	2. Mobilität & Umrüstbarkeit	4.35	Aufgabenteilung	4.60		
			Standardisierung (Arbeitsabläufe/Ansprechpartner:in)	4.52		
			Positionierung von Cobot und Arbeitsmaterial	4.49		
			Ergonomie	4.20		
Unternehmen	(Einmalige) Anschaffungskosten	3.80	Involvement des Managements	4.07	(Fortlaufende) Betriebskosten	4.26
			Kosten der Instandsetzung	3.89		
			Dauer der Instandsetzung	3.65		
			Miteinbeziehung der Gewerkschaften	3.53		

Abbildung 4.3: Übersicht über Erfolgsfaktoren und deren eingeschätzte Relevanz durch Unternehmensvertreter:innen.

Auf Basis der Bewertungen lässt sich zunächst schlussfolgern, dass sich nach Einschätzung der Befragten das Vertrauen in einen Cobot sowohl vor der Einführung ($M = 4.05$) als auch während der Betriebsphase ($M = 4.48$) stark auf den Erfolg einer Cobot-Einführung auswirkt. Wenngleich das dynamische Vertrauen während der Interaktion etwas höher bewertet wurde, wird gleichsam deutlich, dass die befragten Unternehmensvertreter:innen in Einklang mit den dargestellten Vertrauensmodellen (vgl. Kapitel 3.4.7) der Auffassung sind, dass Vertrauensbildungsprozesse bereits vor einer realen Interaktion beginnen und den Erfolg einer Cobot-Einführung bedingen. In der Entscheidungs- und Betriebsphase stellt das initiale bzw. das dynamische Vertrauen jeweils den drittichtigsten Erfolgsfaktor für eine Cobot-Einführung aus Sicht der Unternehmensvertreter:innen dar. In der Entscheidungsphase schrieben die Unternehmensvertreter:innen der grundsätzlichen Eignung des Produktionsprozesses für die Einführung eines Cobots ($M = 4.67$) sowie der Mobilität und Umrüstbarkeit des Cobots ($M = 4.35$) eine höhere Relevanz als dem initialen Vertrauen zu. Dabei handelt es sich um notwendig produktionstechnische Voraussetzungen, ohne die eine Cobot-Einführung grundsätzlich nicht möglich ist.

In der späteren Betriebsphase gelten die Zuverlässigkeit des Cobots ($M = 4.81$) sowie das Vorhandensein etablierter Standards für den Umgang mit Störungen ($M = 4.54$) als relevanteste Faktoren. Dabei stellt die Zuverlässigkeit nach Literaturmeinung wiederum einen entscheidenden Einflussfaktor auf das Vertrauen dar, da eine mangelnde Zuverlässigkeit zu enttäuschten Erwartungen und einem rapiden Vertrauensverlust führen kann. Dieser Faktor ist auch in den meisten Vertrauensmodellen prominent repräsentiert (vgl. Kapitel 3.4.7). Etwas überraschender ist die hohe Bewertung der Standards für Störungen. Hierbei handelt es sich um einen Faktor, dessen beigemessene Relevanz in Abhängigkeit vom Einsatzszenario verhältnismäßig stark variiert. Besonders wenn Unternehmen beabsichtigen, die Flexibilität ihrer Produktion zu erhöhen, erachten die Befragten vordefinierte Standards ($M = 4.84$) als hochrelevant, damit verschiedene beteiligte Personen in verschiedenen Einsatzbereichen mit Störfällen adäquat umgehen können.

Zur Beantwortung der Forschungsfragen stehen besonders die Zusammenhänge zum Anthropomorphismus und zur Angst vor Arbeitsplatzverlust im Vordergrund. Die Unternehmensvertreter:innen schätzten es als hochrelevant ein, dass die Mitarbeitenden nicht befürchten, ihren Arbeitsplatz durch eine etwaige Cobot-Einführung zu verlieren ($M = 4.59$). Dies steht in Einklang zu Erkenntnissen aus Fallstudien, die die außerordentliche Relevanz dieses Faktors nahelegen (vgl. Kapitel 2.3.2). Lediglich der Eindruck der Mitarbeitenden, gut über die anstehenden Veränderungen informiert zu sein ($M = 4.75$), wurde als noch relevanter menschbezogener Faktor in der Implementierungsphase eingeschätzt. Dieser könnte insofern in Bezug zu Ängsten vor Arbeitsplatzverlust stehen, als eine frühzeitige Information diffuse Ängste abbauen kann. Eine entsprechend vorausschauende Kommunikationsstrategie seitens des Managements scheint demnach ein wichtiger Faktor für eine ausreichende Akzeptanz zu sein, die – in Form der Befürwortung einer Einführung

($M = 4.16$) – ebenfalls als wichtig eingestuft wird. Weitere Faktoren, die dazu führen könnten, dass ein Cobot eher als willkommene Erleichterung denn als stresserzeugende Bedrohung im Arbeitsalltag wahrgenommen wird, schätzten die Versuchsteilnehmenden weniger wichtig ein. Die Reduktion der geistigen Beanspruchung der Mitarbeitenden durch den Cobot ($M = 3.25$) sowie die Wahrnehmung des Cobots als Kollegen ($M = 2.91$) wurden nur als *mittelmäßig* relevant beurteilt. Möglicherweise interpretierten die Befragten die Bezeichnung des Cobot als Kollegen als unangemessene Vermenschlichung und Gleichsetzung des Cobots mit menschlichen Kolleg:innen und bewerteten diesen Faktor daher niedrig, wenngleich der Kolleg:innen-Status auch auf eine enge Zusammenarbeit ohne Ängste vor Ersetzung verweisen könnte. Ferner erachteten die Befragten eine an den jeweiligen Mitarbeitenden angepasste Arbeitsgeschwindigkeit des Cobots ($M = 4.15$) als relevant. Eine angepasste Geschwindigkeit galt dabei als geringfügig relevanter als eine möglichst hohe Geschwindigkeit ($M = 4.02$). Dies könnte darauf hindeuten, dass eine Überforderung der Mitarbeitenden beim Arbeitstempo vermieden werden sollte, um den Eindruck zu vermeiden, dem Cobot gegenüber unterlegen zu sein. Allerdings lässt sich bei diesen Faktoren kaum auflösen, ob eher mitarbeiter- oder leistungsbezogene Überlegungen bei der Bewertung im Vordergrund standen, da ein abgestimmtes Arbeitstempo auch in Hinblick auf eine gewünschte flüssige Zusammenarbeit ohne Wartezeiten interpretiert werden könnte.

Im Gegensatz zu den intensiven Forschungsaktivitäten im Bereich des Anthropomorphismus und dessen vermuteten Einfluss auf das Mensch-Cobot-Vertrauen (vgl. Kapitel 3.2.4) wird die Vermenschlichung des Cobots als potenziell vertrauensfördernde Maßnahme nur *mittelmäßig* bis *wenig wichtig* eingeschätzt. Die anthropomorphe Gestaltung des Roboters, sodass dieser einem Menschen möglichst gleicht ($M = 2.15$) oder ihm zumindest ähnelt ($M = 2.26$) sowie die Bezeichnung mit einem menschenähnlichen Namen ($M = 2.31$) sind diejenigen Faktoren, die in der Implementierungsphase als mit Abstand am wenigsten relevant eingeschätzt wurden (gepunktete Umrandung in Abbildung 4.3). Neben diesen eindeutig dem Anthropomorphismus zuordenbaren Faktoren schätzten die Befragten auch weitere Faktoren als vergleichsweise wenig relevant ein, die mit diesem Phänomen in Verbindung gebracht werden können, wie etwa eine gefällige Gestaltung des Cobots ($M = 2.73$) sowie die automatische Erkennung menschlicher Zustände – wie z. B. das Stresslevel der Mitarbeitenden – durch den Cobot ($M = 2.75$). Damit fallen die Relevanzeinschätzungen deutlich geringer aus als bei anderen weichen, mitarbeiterbezogenen Faktoren, die bspw. das Vertrauen oder das Gefühl der Informiertheit auf Seiten der Mitarbeitenden betreffen. Möglicherweise gilt Anthropomorphismus nur als Unterkategorie dieser übergeordneten Faktoren und wird daher nicht als gleichrangig relevant angesehen. Ferner ist zu bedenken, dass die mittels Selbstauskunft erhobenen Relevanzbewertungen ausschließlich expliziten Anthropomorphismus zu erfassen, aber keine impliziten Tendenzen abzubilden vermögen und durch die Aversion gegenüber dem Eingeständnis vermenschlichender Zuschreibungen verzerrt sein könnten (vgl. Kapitel 3.2.3.2).

Eine ähnlich geartete Differenz zwischen expliziten Relevanzeinschätzung und der unbewusst wirkenden Salienz gewisser Faktoren deutet die ergänzende Auswertung der Nennungen von Erfolgsfaktoren in den Freitextfeldern an, die in Abbildung 4.4 ersichtlich ist. Bei dieser Abfrage wurden keine Antwortmöglichkeiten vorgegeben. Die Versuchspersonen nannten überwiegend finanzielle Faktoren, die im weiteren Verlaufe der Studie allerdings nicht zu denjenigen Faktoren gehörten, die als am relevantesten angesehen wurden (einmalige Anschaffungskosten $M = 3.80$, Kosten der Instandsetzung $M = 3.89$, fortlaufende Betriebskosten $M = 4.26$). Die Akzeptanz der Mitarbeitenden als prototypischer menschbezogener Faktor scheint bei der *ad hoc*-Abfrage mit elf Nennungen weitaus weniger präsent zu sein als die finanziellen Faktoren, die einfacher mess-, kalkulier- und gestaltbar sind. Daraus lässt sich folgern, dass Aspekte, die den Unternehmensvertreter:innen schnell in den Sinn kommen und möglicherweise die Debatten um Cobot-Einführungen im betrieblichen Umfeld dominieren, nicht zwangsläufig diejenigen Faktoren widerspiegeln, die als entscheidend für den Erfolg einer Cobot-Einführung erachtet werden.

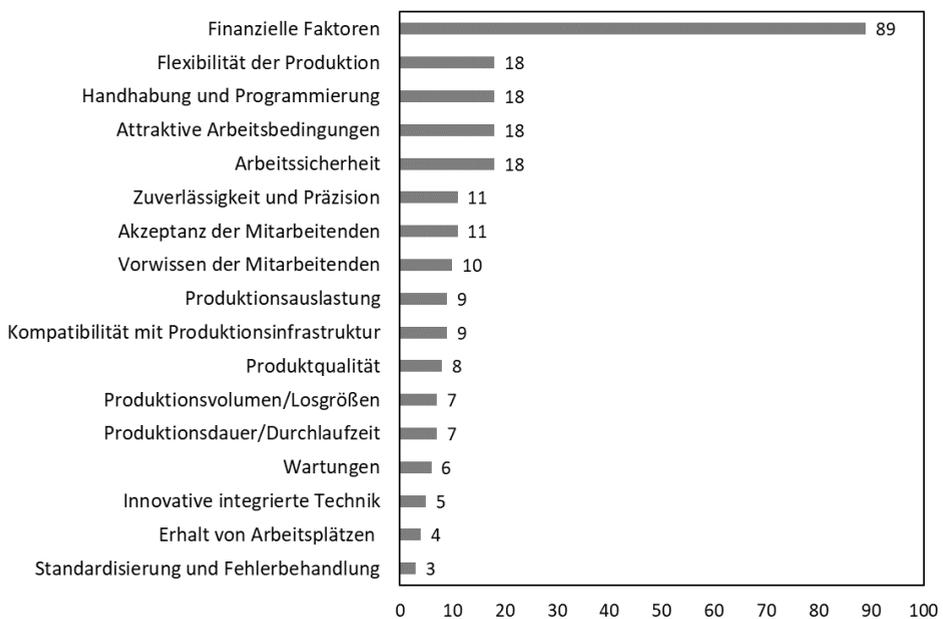


Abbildung 4.4: Anzahl der Nennung von zusätzlichen Erfolgsfaktoren in Freitextfeldern ($n = 251$).

Zusammenfassend lässt sich festhalten, dass sich Parallelen zwischen den Einschätzung der Unternehmensvertreter:innen und der Literaturmeinung zeigen. Neben einer hohen Relevanzeinschätzung für einige technologische und prozessorientierte Faktoren scheint ebenfalls ein Bewusstsein für die entscheidende Rolle weicherer, mitarbeiterbezogener

Faktoren wie Vertrauen und Akzeptanz zu bestehen. Dies gilt bereits für die Entscheidungsphase, bevor die MRI tatsächlich im Unternehmen implementiert ist. Während der Betriebsphase ist in Übereinstimmung mit den gängigsten Vertrauensmodellen die Zuverlässigkeit des Cobots der bestimmende Erfolgsfaktor. Wenig verwunderlich ist, dass die Arbeitssicherheit ($M = 4.83$) insgesamt am höchsten bewertet wurde. Finanzielle Faktoren scheinen zwar kognitiv sehr präsent und salient, aber nicht in gleichem Maße relevant zu sein. Der vermutete starke Einfluss der Angst vor Arbeitsplatzverlust wurde durch die Bewertungen der Unternehmensvertreter:innen bestätigt. In der Forschung thematisierte Determinanten von Vertrauen und Gegenmaßnahmen gegen die Angst vor Arbeitsplatzverlust wie u. a. eine anthropomorphe Gestaltung des Roboters oder eine Reduktion der mentalen Belastung der Mitarbeitenden erhielten im Gegensatz dazu geringe Relevanzwerte. Womöglich reflektieren diese niedrigen Bewertungen – ähnlich wie beim Effekt des unheimlichen Tals – eine ablehnende Haltung, wenn der Cobot begrifflich durch die Umschreibung als Kollege und die Beschreibung dessen Aussehens als menschenähnlich zu stark in die Nähe des Menschen gerückt wird. Ferner ist eine mögliche Verzerrung durch die Selbstauskunft zu berücksichtigen, da sich bspw. impliziter Anthropomorphismus hierdurch nicht erfassen lässt.

Die Diskrepanz zwischen der hohen Aufmerksamkeit in der Forschung und der geringen praktischen Relevanz besonders in Hinblick auf Faktoren, die die menschenähnliche Gestaltung adressieren, lässt sich auf unterschiedliche Art und Weise problematisieren. Aus Sicht der Anwendungsforschung besteht die Gefahr, praktische Bedarfe zu verkennen und sich unzureichend an diesen zu orientieren. Unternehmenspraktiker:innen hingegen könnten die Relevanz möglicher Maßnahmen zur Vertrauenssteigerung unterschätzen. Das würde erklären, warum trotz des Bewusstseins um die Notwendigkeit, die Bedürfnisse und Bedenken der Mitarbeitenden bei Technologieeinführungen einzubeziehen, dies in der Praxis häufig zu selten geschieht und somit den Erfolg solcher Einführungsprozesse gefährdet (Oubari et al., 2018).

Die Ergebnisse bieten eine erste Übersicht über Erfolgsfaktoren und deren eingeschätzte Relevanz in der Praxis. Die rein quantitative Einschätzung der zuvor ermittelten, nicht zwangsläufig vollständigen Liste an Faktoren kann allerdings aufgrund der Methodik nur ein einseitiges Bild vermitteln. Einschränkend ist ferner die nicht-repräsentative Stichprobe zu erwähnen, die nur die Einschätzungen der unternehmerischen Leitungs- und Planungsebene repräsentiert. Um die dahinterstehenden Überlegungen und die Zusammenhänge zwischen den einzelnen Faktoren besser zu durchdringen, wurde in Vorstudie 2 ein vertiefender qualitativer Ansatz gewählt, im Rahmen dessen neben Expert:inneninterviews mit Mitarbeitenden in Leitungsfunktion auch Workshops mit Mitarbeitenden der operativen Ebene durchgeführt wurden.

4.3 Vorstudie 2: Qualitative Studie zu Cobot-Einführungen in der Praxis

Hinweis: Einige Inhalte aus diesem Kapitel sind in Kopp et al. (2020) veröffentlicht.

4.3.1 Fragestellungen und Zielsetzung

Das Ziel der zweiten Vorstudie bestand in einer Überprüfung und Vertiefung der quantitativen Ergebnisse aus der ersten Vorstudie zur Klärung der *FFI*. Demnach liegt der Fokus erneut auf betrieblichen Einführungsprozessen von Cobots. Die zweite Vorstudie verfolgte einen explorativen qualitativen Studienansatz, der auf 17 teilstrukturierten Expert:inneninterviews mit 19 Unternehmensvertreter:innen⁴ und auf vier Workshops mit je sechs bis zehn Fertigungsmitarbeitenden in vier produzierenden KMU mit Sitz in Baden-Württemberg basiert. Die Expert:inneninterviews und Workshops wurden im Rahmen des vom Bundesministerium für Bildung und Forschung (BMBF) geförderten Forschungsprojekts *Proaktive Diagnose und Gestaltung des Cobot-Einsatzes in KMU* (ProBot)⁵ im Zeitraum von März bis April 2019 durchgeführt. Neben dem Verfasser dieser Dissertation wirkten dabei jeweils ein bis zwei weitere Forschende aus dem Projekt mit.

4.3.2 Methoden und Material

4.3.2.1 Expert:inneninterviews

Im Rahmen der Interviews wurden Expert:innen aus vier Unternehmen befragt, die sich den Branchen Landtechnik, Modell- und Formenbau, Metallverarbeitung sowie Elektro-, Blitz- und Überspannungsschutz zuordnen lassen. Alle Unternehmen beschäftigten sich in einer frühen Phase mit der möglichen Einführung von Cobots und hatten dafür potenziell geeignete Anwendungsszenarien in ihren Produktionsprozessen identifiziert. Die Expert:innen für die Interviews entstammten jeweils unterschiedlichen Unternehmensbereichen, sodass sich in Summe ein ganzheitliches Meinungsbild des Unternehmens ergab. Die Unternehmensvertreter:innen verfügten durch ihre spezifische Funktion im Unternehmen über jeweils unterschiedliches Rollenwissen und galten insofern als fachspezifische Expert:innen (Helfferich, 2019, S. 681), wenngleich sie sich mit Cobots nur geringfügig auskannten. Ihr Expert:innen-Status bezog sich demnach nicht auf die betrachtete Technologie, sondern auf die unternehmensinternen Abläufe, Anforderungen und

⁴ Die Diskrepanz zwischen der Anzahl an durchgeführten Interviews und an interviewten Expert:innen ergibt sich daraus, dass zwei Interviews mit je zwei Expert:innen durchgeführt wurden.

⁵ Förderkennzeichen: 02L17C551.

Betrachtungsweisen hinsichtlich Technologieeinführungen. Entsprechend stand bei den Expert:inneninterviews nicht die Gewinnung von Faktenwissen zu Cobots an sich im Vordergrund, sondern von Prozess- und Deutungswissen im Sinne der „Rekonstruktion subjektiver Deutungen und Interpretationen“ (Bogner, Littig & Menz, 2014, S. 2) in Hinblick auf eine Cobot-Einführung im Unternehmen. Die Unternehmensvertreter:innen hatten jeweils eine der folgenden Funktionen inne: kaufmännische / technische Geschäftsführung, technische Leitung / Produktionsleitung / F&E-Leitung, Personalleitung, Arbeitsvorbereitung / Vorarbeiter:in oder Sicherheitsbeauftragte:r.

Alle Expert:innen erklärten im Vorfeld ihr Einverständnis mit der kompletten Audioaufnahme der Interviews und deren anschließender Verwertung im Rahmen wissenschaftlicher Forschung. Sofern die Expert:innen nicht ohnehin bereits firmenseitig in die Überlegungen der Cobot-Einführung eingebunden waren, wurden sie im Vorfeld der Interviews schriftlich und/oder verbal mit dem Begriff des Cobots vertraut gemacht, um ein notwendiges Grundverständnis zu gewährleisten. Die Expert:innen hatten ferner vor Beginn des Interviews die Möglichkeit, Nachfragen zu den einführenden Informationen oder zum Ablauf des Gesprächs zu stellen.

Zur Strukturierung des Gesprächs wurde ein vorab entwickelter teilstandardisierter Interview-Leitfaden (vgl. Anhang B.II.a) eingesetzt, der auf offenen Fragen ohne vorgegebene Antwortmöglichkeiten basiert und typischerweise für halbstrukturierte Interviews eingesetzt wird (Döring & Bortz, 2016, S. 358). Solche leitfadengestützten Expert:inneninterviews gelten als „verbreitete, ausdifferenzierte und methodologisch vergleichsweise gut ausgearbeitete“ (Helfferich, 2019, S. 669) qualitative Datenerhebungsmethode. Der Leitfaden begann mit einer Erzählaufforderung, die notwendige Hintergrundinformationen einholen und eine „flexible und entspannte Interviewsituation“ (Döring & Bortz, 2016, S. 363) schaffen sollte, in der sich die interviewte Person wohlfühlte, sodass sie bereitwillig und unverfälscht die nachfolgenden offen gestellten Fragen beantwortete (Bogner et al., 2014, S. 60f.). Die gewünschten Hintergrundinformationen zielten auf die gegenwärtige Rolle im Unternehmen, aktuelle Themen und Herausforderungen sowie allgemeine Vorerfahrungen mit Industrierobotern und Automatisierungstechnologie ab. Insofern kann der Leitfaden als Kombination der Schemata *Erzählaufforderung-Erzählung* und (überwiegend) *Frage-Antwort* verstanden werden (Helfferich, 2019, S. 675). Der Erstellungsprozess des Leitfadens orientierte sich an dem sog. *SPSS-Prinzip*, das aus den vier sequenziellen Schritten des Sammelns von Fragen, des Prüfens derer Eignung und Relevanz, des Sortierens der verbliebenen Fragen und des Subsummierens derselben zu offenen Leitfragen besteht (Helfferich, 2011, S. 182ff., 2019, S. 677f.).

Das Gespräch fokussierte sich mithilfe offen formulierter Fragen erstens auf grundsätzliche Überlegungen zu Cobots und allgemeine Assoziationen, zweitens auf die unternehmensinternen Auslöser, sich mit Cobots zu befassen, drittens auf die avisierten Ziele der Cobot-

Einführung sowie auf mögliche Vor- und Nachteile dadurch, viertens auf die Anforderungen an die Gestaltung des Einführungsprozesses, fünftens auf konkrete Auswirkungen auf die Beschäftigten, und sechstens auf den jeweiligen vermuteten Mehrwert des Cobot-Einsatzes bei konkreten Arbeitsschritten. Die Themenblöcke adressierten zunächst eher allgemeine Aspekte, die ein weites Feld an Antworten erlaubten. Nach Möglichkeit wurde diese Reihenfolge grob eingehalten, damit zuerst unvoreingenommene allgemeine Assoziationen erhoben werden konnten, bevor bspw. mögliche konkrete Einsatzszenarien diskutiert wurden. Vorrang genoss allerdings das Ziel einer möglichst natürlichen Gesprächsführung (Bogner et al., 2014, S. 29). In diesem Sinne wurden Anregungen oder Bemerkungen des Gegenübers zu anderen Themenblöcken stets aufgegriffen und die Reihenfolge der Themen entsprechend dynamisch angepasst. Ferner wurde der thematische Fokus in der Befragungssituation entsprechend variiert, um den Rollen der Expert:innen im Unternehmen und den damit verbundenen Wissensschwerpunkten gerecht zu werden.

Die Fragen wurden möglichst offen gestellt; Suggestivfragen wurden vermieden. Wenn die Expert:innen zu einer bestimmten Frage keine Antwort wussten oder diese nicht verstanden, konkretisierten die Forschenden die Frage weiter und boten u. U. beispielhafte Antwortmöglichkeiten als Inspiration an. Demnach folgte die Datenerhebung dem Grundprinzip, ein hohes Maß an Offenheit zu gewährleisten und nur wenn nötig zu strukturieren und zu intervenieren (Helfferich, 2019, S. 270). In der späteren Auswertung wurde jeweils differenziert, ob sich die befragte Person entweder ohne weitere Hinweise oder in Folge derartiger Konkretisierungen und Anregungen äußerte, da diese Differenz auf eine unterschiedliche Salienz der Informationen hinweist. Die mittlere Interviewdauer betrug $Mdn = 40:38$ Minuten ($M = 44:18$ Minuten, $SD = 12:47$ Minuten).

Im Nachgang der Interviews wurden sämtliche Aufnahmen volltranskribiert, also im Wortlaut verschriftlicht (Döring & Bortz, 2016, S. 358), um die Aussagen einer qualitativen Inhaltsanalyse unterziehen zu können. Inzwischen existieren viele schwer voneinander abgrenzbare Varianten dieser Analyseverfahren (Schreier, 2014). Der häufig verwendete Ansatz nach Mayring (2015) sieht eine „streng regelgeleitet[e]“ (Mayring & Fenzl, 2019, S. 635) Zuordnung vordefinierter Kategorien vor. Mayring selbst bezeichnet diesen Ansatz als „qualitativ orientierte kategoriengeleitete Textanalyse“ (Mayring & Fenzl, 2019, S. 634). Durch die vorgesehene statistische Analyse der Code-Häufigkeiten enthält dieses Verfahren stets einen quantitativen Analyseschritt. Alternativ basiert die qualitative Inhaltsanalyse nach Gläser und Laudel (2013) auf einer kategoriengeleiteten Reduktion des Materials durch Extraktion, Subsumption und Paraphrasierung, ohne ein exaktes Coding und eine quantitative Auswertung der Codehäufigkeiten zu fordern. Bei dieser sog. *Inhaltsanalyse durch Extraktion* steht im Vordergrund, was gesagt wurde, anstatt wie etwas gesagt wurde (Gläser & Laudel, 2013, S. 30). Durch den Fokus auf die Verdichtung des Primärmaterials entsteht „quasi ein Paralleltext“ (Schreier, 2014, S. 20), der letztlich zum alleinigen Gegenstand der Interpretation wird. Im Gegensatz zur deduktiven Inhaltsanalyse nach

Mayring (2015) favorisiert die Inhaltsanalyse durch Extraktion nach Gläser und Laudel (2013) ein gemischt induktiv-deduktives Vorgehen bei der Kategorienzuordnung, das auch die Generierung neuer Kategorien aus dem Textmaterial heraus erlaubt (Schreier, 2014). Dieses effizientere und stärker qualitativ orientierte Verfahren eignet sich im Rahmen der Vorstudie 2 besser, da Vorstudie 1 bereits mit einem quantitativen Ansatz operierte. Als grobes deduktives Kategoriensystem dienten die für die Erstellung des Interviewleitfadens definierten Themenblöcke. Diese wurden herangezogen, um thematisch gegliederte Exzerpte für jedes Einzelinterview zu verfassen und das Material damit zu verdichten. Weitere thematische Kategorien, die aufgrund des dynamischen Flusses der Expert:inneninterviews *ex ante* nicht vorgesehen waren, wurden bei Bedarf ergänzt.

4.3.2.2 Workshops

Das Ziel der durchgeführten Workshops bestand darin, Einblicke in die Sichtweise derjenigen Mitarbeitenden zu erlangen, die potenziell künftig direkt mit einem Cobot zusammenarbeiten werden. Die Teilnehmenden der Workshops wurden unternehmensseitig auf Basis von Auswahlhilfen festgelegt, die von der Workshop-Leitung bereitgestellt worden waren. Darin wurde darauf hingewiesen, dass ein breites Spektrum an Mitarbeitenden mit unterschiedlicher Technikaffinität und Voreinstellung gegenüber Robotern wünschenswert ist, um das mutmaßlich heterogene Meinungsbild innerhalb der Firmen möglichst ganzheitlich einzufangen. Alle Workshop-Teilnehmenden kamen aus Unternehmenssicht für die künftige Arbeit mit einem Cobot infrage und umfassten z. B. Werker:innen in der manuellen Montage, im Werkzeugbau oder in der Maschinenwartung. Im Rahmen des Workshops wurden positive und negative Assoziationen der Mitarbeitenden sowie deren Anforderungen an einen Cobot und an dessen Einführungsprozess formuliert und auf einer Wichtigkeitsskala eingeordnet. Eine genaue Beschreibung des Workshop-Ablaufs befindet sich in Anhang B.II.b.

Zur Ergebnissicherung wurden die Skalen mit aufgeklebten Notizzetteln fotografisch dokumentiert, wie Abbildung 4.5 und Abbildung 4.6 zeigen. Die beteiligten Unternehmen erhielten im Nachgang ein aufgearbeitetes Dokument mit den Workshop-Ergebnissen und einer schriftlichen Gesamteinschätzung durch die Workshop-Leitung.

subjektiven Blick auf das Material zu reflektieren und ein ausreichendes Maß an Inter-subjektivität zu gewährleisten (Kuckartz, 2018, S. 46). Bei der Interpretation wurde der spezifische und in der qualitativen Forschung unhintergehbare Rollenkontext der befragten Person und die daraus entstehenden Spezifika der Befragungssituation zwischen Forschenden und Workshop-Teilnehmenden berücksichtigt (Helfferich, 2019, S. 683f.). Die Eindrücke aus den Workshops rundeten das detaillierte Bild aus den Interviews ab und ergänzten es um die Sichtweise der operativ unmittelbar betroffenen Mitarbeitenden. Die Ergebnisse werden im Folgenden durch besonders plakative Originalzitate der interviewten Unternehmensvertreter:innen ergänzt. Das Kürzel am Zitatende kennzeichnet dabei jeweils das Unternehmen und die befragte Person.

4.3.3 Ergebnisse und Diskussion

Bei den Befragungen offenbarte sich, dass die Assoziationen zu Cobots zu einem großen Teil dem bekannten Wissen zu Automatisierungsprozessen und Industrierobotern entstammten. Offenbar aktivierte die Vorstellung eines Cobots das mentale Modell von einer Maschine oder konkreter von einem klassischen Industrieroboter. Dies spiegelte sich u. a. in der Vorstellung wider, ein Cobot könne und solle im Gegensatz zu den Mitarbeitenden rund um die Uhr eingesetzt werden:

„Ein Roboter muss laufen. 24-7 am besten. Alles andere ist eigentlich Verschwendung von Geld. Wenn man das nicht schafft, annähernd zu erreichen, ich glaube, dann lässt man es besser.“ [U1E3]

„Der kann 24 Stunden arbeiten, sieben Tage und so weiter. Ich sehe im Moment, ganz ehrlich, keinen echten Zeitgewinn in Zusammenarbeit mit dem Menschen.“ [U4E3]

Auf konkrete Nachfrage waren die befragten Expert:innen in der Lage, dies entsprechend zu reflektieren:

„Ja, vielleicht denke ich da noch nicht so wirklich kollaborierend, sondern schon auch noch eher klassisch.“ [U1E3]

„Wir denken ja aktuell immer noch in dieser Automatisierung, aber vielleicht ist ja genau dieses Zusammenspiel die beste [Lösung] – oder es wird sicherlich bestimmt öfters die beste Lösung sein.“ [U1E2]

Erwartete positive Auswirkungen des möglichen Cobot-Einsatzes orientierten sich ebenfalls häufig an klassischen Produktivitäts- und Qualitätskennzahlen, die insbesondere

durch das Ausbleiben von durch die Mitarbeitenden bedingten Schwankungen verbessert werden sollen:

„[A]uf Qualitätskennzahlen hat er bestimmt einen Vorteil, der macht keine Fehler. Der macht immer die gleiche Tätigkeit und die macht er sehr präzise. Also hat er auf jeden Fall Vorteile auf betriebliche Kennzahlen. [Ein Cobot] ist nicht krank, es ist vieles besser planbar, ja gut, er kann ausfallen. Aber es ist schon besser planbar wie ein Mitarbeiter, der halt auch mal einen guten Tag und einen schlechten Tag hat. Einen Tag mit guter Produktivität und einen Tag mit schlechter Produktivität. Hat alles Vorteile auf die Kennzahlen, klar.“
[U1E3]

Das entsprechende differenzierende und namensgebende Merkmal der Kollaborationsfähigkeit wurde in diesem mentalen Modell nicht repräsentiert, obgleich es auf Nachfrage mit wünschenswerten Effekten wie zusätzlichen Einsatzmöglichkeiten sowie Vorteilen in der Mitarbeitendenakzeptanz assoziiert wurde. Dieser bis dato ungewohnte Handlungsspielraum schien sich nur schwer in die bestehenden mentalen Modelle der Expert:innen integrieren zu lassen. Die Befragten taten sich schwer, sich von den gewohnten Denkmustern zu lösen, zumal die Vorstellung einer sprichwörtlichen Hand-in-Hand-Zusammenarbeit erfordert, die Relation zwischen Werker:in und Roboter sowie das Verständnis des Roboters als reines Werkzeug zu hinterfragen. Entsprechend benannte keiner der Expert:innen die Kollaborationsfähigkeit als Vorteil eines Cobots oder als Anforderung für dessen Einsatz. Auch in den Workshops schärfte sich erst im Laufe der Zeit das Bewusstsein für interaktive Anwendungsmöglichkeiten. Dabei blieb eine gewisse Skepsis unter den Mitarbeitenden erhalten, ob ein Cobot tatsächlich sein Leistungsversprechen erfüllen und zuverlässig genug funktionieren würde. Großes Vertrauen wurde der Sicherheit des Cobots entgegengebracht, der demnach kein Risiko für Arbeitsunfälle und Verletzungen darstelle.

In der Praxis könnten womöglich koexistente und kooperative Anwendungsfälle für den Einstieg in die Mensch-Cobot-Interaktion geeignet sein, um anschließend das volle Potenzial von Cobots in echt-kollaborierenden Anwendungsfällen zu erschließen. So wünschte sich ein Experte, „einen Weg zu finden (...) [hin zu] den echten Kollaborationen“, da er dort „langfristig (...) ganz deutlich“ [U3E4] das größere Potenzial sehe. Aus der Erfahrung mit früheren Technologieeinführungen seien zu Beginn eines solchen Prozesses noch nicht alle Einsatzmöglichkeiten absehbar:

„[A]ber aus der Erfahrung heraus weiß ich schon, wenn mal sowas da ist und man hat neue Möglichkeiten, ergeben sich schlagartig viele Ideen, die wir heute eigentlich gar nicht abschätzen können oder sehen können oder begreifen können.“ [U1E2]

Sowohl von Seite der interviewten Expert:innen als auch der Workshop-Teilnehmenden wurde die Angst vor Arbeitsplatzverlust häufig thematisiert. Sie ist nach Meinung beider Personengruppen eine entscheidende Ursache für Bedenken und Skepsis gegenüber Technologieeinführungen, wenngleich mit Cobots vergleichsweise weniger Ängste assoziiert wurden als mit konventioneller Automatisierung. Bei der kollaborierenden Anwendung seien Mensch und Maschine „zusammen unterwegs“ [U3E4], wohingegen bei nicht-kollaborierenden Anwendungsfällen „der Roboter ja schon in einer gewissen Autonomie arbeitet, was ja näher an der klassischen Automatisierung dran ist“ [U3E4]. Die Akzeptanz ließe sich auch dadurch befördern, dass ein „deutlich qualitativer Unterschied in der Tätigkeit“ [U3E1] des Mitarbeitenden bzw. des Cobots und damit weiterhin ein Gefühl der Sicherheit und Wertschätzung gewährleistet werde. Insgesamt fiel bei den durchgeführten Erhebungen auf, dass häufiger über Akzeptanz denn über Vertrauen diskutiert wurde. Im betrieblichen Arbeitsplatzumfeld scheint der Begriff der Akzeptanz deutlich gängiger und besser fassbar zu sein, wenngleich er in der Beschreibung häufig sehr ähnliche Facetten wie Vertrauen aufwies.

In Bezug auf die genannten Ängste vor Arbeitsplatzverlust attestierten die Befragten eine hohe Eigendynamik im sog. Flurfunk und in der sprichwörtlichen Gerüchteküche. So berichtete ein Experte, dass die Mitarbeitenden bspw. befürchteten, dass die Einführung von Cobots nur der Beginn einer Automatisierungswelle sein könnte, die ihre Arbeitsplätze gefährde:

„Wenn er [der Cobot] diese Tätigkeit übernehmen kann, könnte er vielleicht auch meine übernehmen.“ [U3E1]

Sogar getätigte Investitionen in die langfristige Erhaltung eines konkreten Unternehmensstandorts waren von den Mitarbeitenden in der Vergangenheit bereits als Hinweis darauf gedeutet worden, dass ein Verkauf der Firma und damit einhergehend die Entlassung großer Teile der Belegschaft bevorstehe. Außerdem sei es der Akzeptanz abträglich, wenn sich Mitarbeitende „getrieben“, „nicht wertgeschätzt“ [U3E1] oder „unter Druck“ [U4E3] fühlten, weil der Cobot zu schnell arbeite und damit das Gefühl bei den Mitarbeitenden erzeuge, nicht mithalten zu können. Auch in diesen Aussagen lässt sich erkennen, wie sensibel die Mitarbeitenden auf Veränderungen in der beschriebenen Beziehung und Hierarchie zum Cobot reagieren. Eine ablehnende Haltung sei insbesondere dann zu erwarten, wenn der Cobot den Arbeitstakt für die Mitarbeitenden vorgebe.

Daneben bestanden allgemeiner gefasste und häufig subtilere Ängste in Bezug auf durch Technologieeinführungen bedingte Veränderungen im Produktionsablauf. In diesem Zusammenhang wurde die jüngere Belegschaft als weniger kritisch gegenüber Veränderungen sowie als interessierter und experimentierfreudiger in Bezug auf neue Technologien beschrieben. Grundsätzlich falle die Technologieakzeptanz allerdings personenindividuell

höchst unterschiedlich aus. Das deckt sich mit Erkenntnissen aus anderen Studien, die darauf hinweisen, dass das Alter der Beschäftigten häufig mit anderen Variablen wie der Technikkompetenzüberzeugung und praktischen Vorerfahrungen mit Technologie konfundiert, die letztlich die Einstellung gegenüber Robotern deutlich stärker beeinflussen (Müller-Abdelrazeq, 2020, S. 38f.). Die Technikkompetenzüberzeugung beschreibt dabei eine subjektive Erwartung von Handlungsmöglichkeiten in technikrelevanten Situationen, beeinflusst durch kumulierte Erfahrungen mit vertrauter Technik und durch die Anpassungsfähigkeit an noch unbekannte Technologien (Neyer, Felber & Gebhardt, 2012, S. 88). Gewisse Ressentiments müssten einkalkuliert werden, die sich zwar bei vielen Mitarbeitenden bspw. durch transparente Kommunikation abbauen lassen. Gleichwohl sei zu akzeptieren, dass einige Beschäftigte auf ihrer skeptischen Haltung beharrten. Insbesondere wurde in diesem Zusammenhang der in der Forschung häufig vertretene Annahme widersprochen, dass alle Mitarbeitenden es als Aufwertung und Entlastung erleben, wenn ihre Tätigkeiten vielseitiger und weniger monoton ausgestaltet werden:

„Es gibt ja Mitarbeiter, die sind so gestrickt, die machen am liebsten immer nur immer das gleiche. [...] Möglichst nicht viel denken, sondern hier einfach nur das und das und das machen.“ [U1E3]

In Hinblick auf akzeptanzförderliche Faktoren bestand ein breiter Konsens in den Expert:inneninterviews und Workshops, dass die rechtzeitige, transparente und umfassende Kommunikation über eine bevorstehende Cobot-Einführung gegenüber den Mitarbeitenden von entscheidender Bedeutung für deren Erfolg ist. Die Kommunikation sollte die Benennung der verfolgten Ziele und des avisierten Nutzens durch den Cobot-Einsatz für das Unternehmen sowie für dessen Mitarbeitende beinhalten und auch konkrete Vorteile wie z. B. einen Beitrag zur Standortsicherung darstellen. Ferner sei es im MRI-Kontext wichtig, den Charakter der Zusammenarbeit zu betonen, um das eingetübte Mensch-Technik-Verhältnis aufzubrechen und auf die intensivere Interaktion hinzuweisen. Konkrete Erfolgsgeschichten (*success stories*) aus vergleichbaren Firmen könnten zur Glaubwürdigkeit der Zielbilder beitragen. Ferner wurde darauf hingewiesen, dass mittels der Unternehmenskommunikation ein gezieltes Erwartungsmanagement betrieben werden sollte. Dem Zeitpunkt der Informationsvermittlung komme dabei eine wichtige Rolle zu. So könne eine sehr frühzeitige Kommunikation über eine anstehende Cobot-Einführung die Erwartung erwecken, dass dieser zeitnah in der Produktion anzutreffen sei. Ein zu starker Zeitversatz würde dann möglicherweise als Indiz gewertet, dass das Projekt nicht vorankomme und zum Scheitern verurteilt sei.

Seitens der Belegschaft wurde ein partizipativer Ansatz, bei dem die Ideen der Mitarbeitenden aktiv in den Planungsprozess einbezogen werden, als akzeptanzsteigernd gewertet. Ebenfalls besonders häufig wurde darauf verwiesen, dass Berührungspunkte und praktische Erfahrungen im Umgang mit dem Cobot wichtig seien und sich wesentlich positiver

auf die Akzeptanz auswirkten als eine rein theoretische Beschäftigung. Ein Demonstrator in der Firma wurde als Möglichkeit genannt, insbesondere um erste als wichtig empfundene positive Erfahrungen im Umgang mit einem Cobot zu ermöglichen. Entsprechend sollten gerade in den ersten Tagen der Betriebsphase möglichst keine Funktionsstörungen auftreten. Das schließt an die beschriebene asymmetrische Vertrauensentwicklung an, bei der gerade in einer frühen Phase situativ enttäuschte Erwartungen zu starken Minderungen des Vertrauensniveaus führen können (vgl. Kapitel 3.4.1). Darüber hinaus wünschten sich die Workshop-Teilnehmenden Freiräume zum Experimentieren mit dem Cobot. Das stand in Einklang mit der Ansicht der Management-Ebene, dass die Mitarbeitenden über ein hohes Kreativpotenzial im Umgang mit dem Cobot verfügen und sich schnell an neue Anwendungen wagen würden. Aus anfänglichem ziellosem Herumprobieren könnten sich auf diese Weise rasch nützliche Ideen für neue Anwendungsfälle oder Umsetzungsvarianten in der Produktion ergeben.

Das anthropomorphe Design von Robotern wurde weder von Seiten der Belegschaft noch von Management-Seite thematisiert. Gleichwohl fielen in den Workshops einige vermenschlichende Aussagen und zahlreiche Vergleiche mit menschlichen Kolleg:innen auf. So wurde spontan – wenngleich mit scherzhaftem Tonfall – als erste Anforderung an einen Cobot genannt, dass dieser „blond sein solle“, was sich als Form einer schwachen Vermenschlichung interpretieren lässt. Zwar fehle bei der Interaktion mit einem Cobot der zwischenmenschliche Kontakt, allerdings biete dieser auch den Vorteil, dass er „mache, was man wolle“, ohne dass man mit ihm diskutieren müsse, wie es mit menschlichen Kolleg:innen der Fall sei. Zwar handelt es sich bei dieser Aussage um keine Vermenschlichung i. e. S., doch ist auffällig, dass damit ein direkter Vergleich zwischen zwei Arten von Kooperationspartner:innen gezogen wird, nämlich zwischen bekannten menschlichen Kolleg:innen und neuen Kooperationspartner:innen in Form von Cobots. Menschen und Cobots scheinen durch diese Verortung auf einer Vergleichsebene als in gewisser Weise kommensurabel betrachtet zu werden. Als *tertium comparationis* wurde der Diskussionswille herangezogen. Zwar sprach der Workshop-Teilnehmende dem Cobot einen expliziten Willen und die Fähigkeit, Widerstand gegen Anordnungen zu leisten, explizit ab. Allerdings konzidierte er durch die Äußerung gleichsam, dass er die prinzipielle Möglichkeit bedacht hat und für bedenkenswert hält. In Bezug auf ein einfaches Werkzeug wäre ein Verweis auf einen nicht vorhandenen Diskussionswille vermutlich nicht notwendig gewesen. Wenngleich also keine Vermenschlichung i. e. S. vorliegt, weisen einige Äußerungen dennoch darauf hin, dass sowohl mentale Modelle von klassischen Industrierobotern als auch von Menschen zur differenzierenden Beurteilung des Cobots in Anwendung gebracht wurden.

Ferner wurde die Frage nach der Bewertung einer anthropomorphen Gestaltung und einer Ausstattung mit sozialen Funktionen kontrovers diskutiert. Während dies intuitiv zumeist als unnötig zurückgewiesen wurde, äußerten sich einige Workshop-Teilnehmende im

Zeitverlauf zunehmend positiver zu solchen Funktionen. So bekundeten manche Teilnehmenden, dass sie sich z. B. über eine persönliche Begrüßung durch einen Cobot freuen würden. In der Gesamtschau deuten die Beobachtungen auf eine Diskrepanz zwischen implizitem und explizitem Anthropomorphismus hin. Während explizite Nachfragen mitunter reflexartig zurückgewiesen wurden, zeigten sich bei genauer Betrachtung einiger Aussagen durchaus implizite Formen des schwachen Anthropomorphismus und Denkfiguren, die mentalen Modelle von Menschen entstammten. Vor diesem Hintergrund scheint es lohnend, das Phänomen des Anthropomorphismus differenziert zu betrachten, ohne sich dabei ausschließlich auf *ad hoc*-Selbstausskünfte zu verlassen.

4.4 Zwischenfazit

Die Vorstudien verfolgten das Ziel, die Rolle des Mensch-Cobot-Vertrauens am industriellen Arbeitsplatz auch vor dem Hintergrund potenziell vertrauensfördernder Maßnahmen aus der Sicht von Unternehmensvertreter:innen zu beleuchten. Grundsätzlich ergibt sich aus den beiden Vorstudien, die auf unterschiedlichen Forschungsparadigmen basierten, ein zum Teil konvergentes und zum Teil komplementäres Bild (vgl. Kelle, 2019, S. 163). Vertrauen wurde in der quantitativen Vorstudie 1 als sehr relevant eingeschätzt, und zwar in erster Linie im Verlauf einer MRI, aber auch bereits in deren zeitlichem Vorfeld. Abgesehen von klassischen produktionstechnischen, sicherheitsbezogenen und finanziellen Faktoren, die besonders salient, aber nicht in gleichem Maße relevant zu sein scheinen, besteht ein Bewusstsein für die Relevanz menschenbezogener Faktoren. In der qualitativen Vorstudie 2 zeigte sich ein stark unterschiedliches Vertrauen bezüglich verschiedener Sachverhalte. Während die Mitarbeitenden sehr stark in die Sicherheit eines Cobots vertrauten, herrschte gegenüber der Zuverlässigkeit und Leistungsfähigkeit eher Skepsis vor. Besonders starke Bedenken offenbarten sich in Hinblick auf den möglichen Verlust des eigenen Arbeitsplatzes durch eine technologische Ersetzung. Diese Bedenken wurden *unisono* als maßgebliches Hemmnis für den Erfolg von Technologieeinführungen betrachtet. Entsprechend zeigte sich in der qualitativen Vorstudie 2, dass die Mitarbeitenden sehr genau beobachteten, wie der Cobot, dessen Einsatzzweck und ihre Relation zum Cobot umschrieben wurde. Dies unterstreicht die Erkenntnisse aus früheren Studien, dass nur ein schmaler Grat zwischen der Wahrnehmung als Unterstützung und als Konkurrent verläuft (Weiss et al., 2016) und bereits feine sprachliche Variationen beeinflussen können, wie Menschen einen Cobot wahrnehmen (Brandstetter, 2017; Dijksterhuis & Bargh, 2001). Laut der Meinung der interviewten Expert:innen und Workshop-Teilnehmenden lassen sich diese existenziellen Ängste auch dann nicht bei allen Mitarbeitenden beseitigen, wenn eine transparente und partizipative Kommunikationsstrategie verfolgt wird, die in beiden Vorstudien als wünschenswert und relevant eingeschätzt wurde. Das liegt daran, dass gängige Narrative von Cobots als Entlastung von monotonen Aufgaben nicht bei allen Mitarbeitenden

verfangen, zumal die Ansprüche an die eigene Arbeit stark heterogen ausfallen. Manche Mitarbeitende schätzen gerade solche Aufgaben besonders, die aus der Perspektive der Forschung aufgrund ihrer Monotonie als belastend bewertet werden.

Die in Vorstudie 2 erkennbaren Schwierigkeiten, sich in die Kollaboration zwischen den Mitarbeitenden und dem Cobot hineinzudenken, weisen auf einen latenten Einfluss bestehender mentaler Modelle hin, die der bekannten Automatisierungs- und Rationalisierungslogik folgen. Erst nach der Verdeutlichung der Differenz zwischen konventionellen Robotern und Cobots wurden letztere als akzeptanzfähigere Technologie mit verändertem produktionsbezogenem Potenzial eingestuft, bei der die Mitarbeitenden am Produktionsablauf direkt beteiligt bleiben können. Das gilt allerdings nur unter der Voraussetzung, dass die Absichten der Unternehmensleitung beim Einsatz dieser Cobots gerade nicht aus einer klassischen Automatisierungslogik resultieren. Für die Relevanz von Anthropomorphisierung zeigten sich allenfalls implizite Hinweise durch vermenschlichende Aussagen in Vorstudie 2, wohingegen dieser Thematik bei expliziten Fragen in beiden Vorstudien nur eine geringe Relevanz beigemessen wurde. Demnach scheint diese Thematik in der Forschung eine deutlich höhere Aufmerksamkeit als in der unternehmerischen Praxis zu erfahren. Allerdings offenbarten die Diskussionen in den Workshops einige Kontroversen und interindividuell stark unterschiedliche Meinungen, die mitunter auch durch eine starke Ablehnung jeglicher Menschenähnlichkeit gekennzeichnet waren. Ein homogeneres Meinungsbild, das sich mit Erkenntnissen aus der Forschungsliteratur deckt, ergab sich hingegen in Hinblick auf den positiven Einfluss von Vorkenntnissen und insbesondere von praktischen Vorerfahrungen mit Robotern auf das Vertrauen. Ob die Möglichkeit, praktische Erfahrungen zu erwerben, tatsächlich zu einem erhöhten Vertrauen in einen Cobot führen würde, lässt sich gleichwohl anhand des verwendeten Forschungsdesigns basierend auf antizipativen Selbstauskünften nicht abschließend beantworten.

Der experimentelle *mixed methods*-Ansatz eignete sich gut, um einerseits einen empirisch breit gestützten Überblick über mögliche Erfolgsfaktoren sowie die diesen beigemessene Wichtigkeit zu erlangen und andererseits um diese Erkenntnisse im qualitativen Design auf Basis der verschiedenen Meinungsbilder von Vertreter:innen aus leitender und operativer Ebene zu triangulieren und zu vertiefen. Daraus ergibt sich in Summe ein kohärentes, valides und vertieftes Gesamtbild. Nichtsdestotrotz ist einschränkend zu erwähnen, dass es sich weder um repräsentative noch um – insbesondere in Hinblick auf die qualitative Erhebung – objektive oder experimentell bestätigte Erkenntnisse handelt.

4.5 Experimentalstudie 1: Online-Studie mit Produktionsmitarbeitenden

Hinweis: Einige Inhalte aus diesem Kapitel (inklusive einiger Abbildungen) sind in Kopp et al. (2022) veröffentlicht.

4.5.1 Fragestellungen und Studiendesign

Basierend auf der Herleitung in Kapitel 3.5 verfolgt die erste Experimentalstudie das Ziel, die nachfolgenden Forschungsfragen zu klären:

FF2: Inwiefern lässt sich die Wahrnehmung eines Cobots und der Mensch-Cobot-Relation mittels sprachlichen Framings beeinflussen?

FF3: Wie wirken sich die wahrgenommene Menschenähnlichkeit eines Cobots und die wahrgenommene Kooperationsorientierung der Mensch-Cobot-Relation auf das Mensch-Cobot-Vertrauen aus?

Abbildung 4.7 illustriert schematisch die relevanten theoretischen Zusammenhänge und die daraus resultierende experimentelle Idee, die der Experimentalstudie zugrunde liegt.

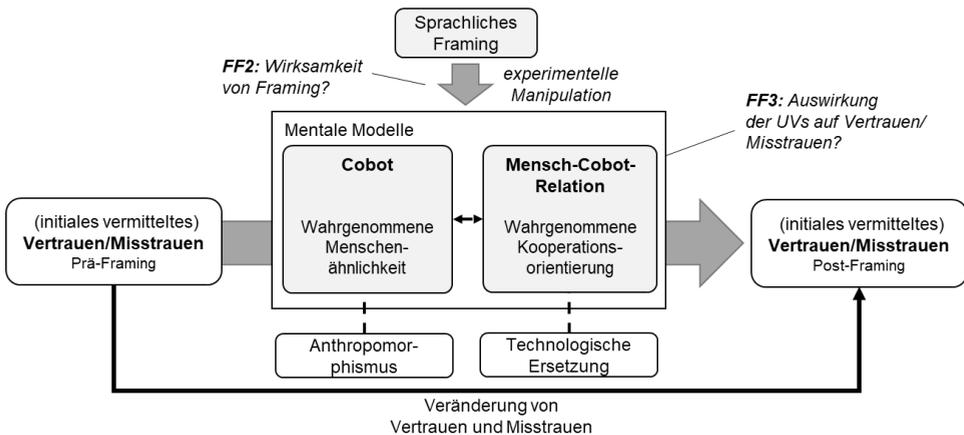


Abbildung 4.7: Schematische Darstellung des Designs von Experimentalstudie 1.

Die wahrgenommene Menschenähnlichkeit und die wahrgenommene Kooperationsorientierung der Mensch-Cobot-Relation operationalisieren die zentralen Aspekte der mentalen Modelle und stellen die beiden UVs dar. Das sprachliche Framing wird genutzt, um diese

Variablen durch unterschiedliche Cobot-Beschreibungen experimentell zu manipulieren, was die Durchführung der Studie mit verschiedenen Versuchsgruppen erfordert. Der betreffende Cobot wird in diesen Texten entweder menschenähnlich oder maschinell und die Mensch-Cobot-Relation entweder kooperations- oder konkurrenzorientiert beschrieben. Ein kooperierendes⁶ Verhältnis ist dadurch gekennzeichnet, dass Mitarbeitende den Cobot als nützliche Unterstützung und als komplementäre Ergänzung betrachten und sich aus dessen Nutzung Vorteile versprechen. Wenn Mitarbeitende den Cobot allerdings als potenzielle Bedrohung für ihren Arbeitsplatz und/oder als stresserzeugende Konkurrenz innerhalb eines Wettstreits um die höhere Leistungsfähigkeit empfinden, lässt sich von einem konkurrierenden Verhältnis sprechen. Die jeweilige Repräsentation der Mensch-Cobot-Relation hängt demnach davon ab, wie der Mitarbeitende sich und seine Fähigkeiten relativ zum Cobot und dessen vermuteten Fähigkeiten verortet und in welchem organisationalen Kontext die Interaktion stattfindet.

Die Ergebnisse der Manipulationskontrolle liefern Einsichten in die Wirksamkeit des Framings (*FF2*), indem sie offenlegen, inwiefern sich die wahrgenommene Menschenähnlichkeit und die wahrgenommene Kooperationsorientierung durch die experimentelle Manipulation verändert haben. Anschließend wird analysiert, wie diese herbeigeführten Unterschiede in der Beschaffenheit der mentalen Modelle das initiale Vertrauens- und Misstrauensniveau der Mitarbeitenden in Bezug auf den Cobot beeinflussen (*FF3*). Dazu wird das Vertrauens- und Misstrauensniveau jeweils vor und nach dem Framing gemessen, sodass letztlich ein Prä-Post-Versuchsdesign mit der je Versuchsgruppe unterschiedlichen experimentellen Manipulation der zwei UVs als zweistufige Zwischensubjektfaktoren vorliegt. In Kombination mit einer zusätzlichen Kontrollgruppe, die keine experimentelle Manipulation erhielt, ergeben sich somit fünf unterschiedliche Versuchsbedingungen.

4.5.2 Methoden und Material

4.5.2.1 Stichprobenbeschreibung

Die Teilnehmenden an der als Online-Studie konzipierten Experimentalstudie 1 wurden mithilfe des Panel-Anbieters *ToLuna Deutschland GmbH* rekrutiert und mussten gewisse Teilnahmevoraussetzungen erfüllen. Das Ziel bei der Auswahl bestand darin, möglichst solche Personen zu adressieren, die zwar noch nicht über praktische Erfahrungen im Umgang mit Robotern verfügten, aber aufgrund ihrer Tätigkeit künftig am ehesten mit Cobots konfrontiert werden könnten. Daher mussten alle Teilnehmenden aktuell im Bereich der

⁶ Der Begriff *kooperierend* ist hier im allgemeinen Sprachgebrauch gemeint und bezieht sich nicht auf die Kooperation als spezifische Unterart der MRI, wie in Kapitel 2.1.2 dargestellt. Kooperation und Konkurrenz werden als gegensätzliche Strategien, sich zu möglichen Interaktionspartnern ins Verhältnis zu setzen, in verschiedenen Kontexten diskutiert, siehe u. a. Tauer und Harackiewicz (2004).

Produktion oder der Fertigung beschäftigt sein, dort einer operativen Tätigkeit nachgehen und durften über keine praktischen Erfahrungen in der Arbeit mit Robotern verfügen. Der Mangel an Vorerfahrungen und an damit einhergehenden stabilen mentalen Modellen von Robotern wurde als begünstigendes Kriterium für die Effektivität des Framings vermutet.

Insgesamt erfüllten 256 Personen die Teilnahmevoraussetzungen. Davon wurden im Verlaufe der Befragung 60 Teilnehmende ausgeschlossen, weil sie die Kontrollfragen zum Verständnis der zuvor präsentierten Texte nicht korrekt beantworteten. Der mit 76.6 % hohe Anteil richtiger Antworten im Vergleich zur Ratewahrscheinlichkeit von 5 % bei den Kontrollfragen lässt vermuten, dass die verbleibenden 196 Versuchspersonen die Texte tatsächlich gründlich gelesen hatten. 18 Teilnehmende brachen im weiteren Verlauf die Studie ab, sodass 178 vollständige Datensätze zur Verfügung standen.

Im Rahmen der Datenbereinigung nach Abschluss des Erhebungszeitraums wurden sog. *straightliner*, die nach einheitlichen Antwortmustern vorgingen (Koch, Peter & Müller, 2019, S. 160f.), sowie Durchklicker (*speeder*), deren geringe Verweildauer eine flüchtige und wenig sorgfältige Bearbeitung der Fragen vermuten ließ (Prill, 2017, S. 229), aus dem Datensatz entfernt. Als *straightliner* galten diejenigen Teilnehmenden, die in mindestens einer Multi-Item-Skala mit mehr als vier Items jeweils immer den gleichen Skalenwert auswählten. Als Durchklicker galten ferner alle Teilnehmenden, die weniger als die Hälfte des Medians der Gesamtbearbeitungszeit zur Bearbeitung benötigten (Theobald, 2017, S. 75f.). Durch diese beiden Maßnahmen wurden 29 Datensätze aus der Datenbasis entfernt. Letztlich standen demnach 149 gültige und qualitätsgesicherte Datensätze für die Auswertung zur Verfügung. Vier Datensätze (davon drei aus der Kontrollgruppe) wurden anhand von Box-Plots als Ausreißer identifiziert und von der Analyse ausgeschlossen. Die folgenden Auswertungen basieren daher auf einer Stichprobe von $n = 145$ Datensätzen.

Das Alter der Studienteilnehmenden lag zwischen 18 und 69 Jahren ($M = 47.06$, $SD = 12.06$). 51 Teilnehmende waren weiblich (35.2 %) und 94 männlich (64.8 %). Sie verfügten über eine durchschnittliche Berufserfahrung von $M = 16.16$ Jahren ($Min = 1$, $Max = 47$, $SD = 12.11$) und stimmten im Durchschnitt *teils-teils* bzw. *überwiegend* zu ($M = 3.57$, $SD = 0.86$), dass ihnen der Umgang mit neuen Technologien leichtfällt. Im Mittel hatten sich die Teilnehmenden im Vorfeld *eher weniger* bis *mittelmäßig* intensiv mit Robotern auseinandergesetzt ($M = 2.41$, $SD = 1.00$), bspw. durch Filme, Zeitungsartikel oder im Internet. Außerdem zeigten die Teilnehmenden eine *mittelmäßig* negative Grundeinstellung gegenüber Robotern ($M = 2.91$, $SD = 0.62$).

Alle Umfrageteilnehmenden arbeiteten in der Produktion in einer der folgenden Wirtschaftszweige: Verarbeitende/produzierende Industrie (63.4 %), Baugewerbe (10.3 %), Gesundheits- und Sozialwesen (5.5 %), Handel (2.8 %), Transport und Lagerung (2.8 %), Erbringung von freiberuflichen, wissenschaftlichen und technischen Dienstleistungen

(2.1 %) oder in einem sonstigen Wirtschaftszweig (13.1 %). Der überwiegende Anteil der verarbeitenden Industrie erklärt sich dadurch, dass eine Anstellung in der Produktion/Fertigung zu den Teilnahmevoraussetzungen zählte.

4.5.2.2 Material

Als Stimuli für die experimentelle Manipulation wurden unterschiedlich geframte Texte verwendet. Alle Versuchsteilnehmenden wurden zunächst gebeten, sich vorzustellen, dass sie an ihrem Arbeitsplatz künftig mit einem Cobot zusammenarbeiten werden (*cover story*). Je nach Versuchsbedingung variierte die Beschreibung des Cobots in Hinblick auf dessen Menschenähnlichkeit und die Beschreibung der Mensch-Cobot-Relation in Hinblick auf deren Kooperationsorientierung. Gemäß der Klassifikation in Kapitel 3.3.2 kamen in den jeweiligen Beschreibungen unterschiedliche Hervorhebungsframes zum Einsatz. Die Texte in der Kontrollbedingung enthielten keine Informationen über Cobots, sondern über Just-in-time-Produktion als willkürlich gewähltes Alternativthema im Produktionskontext. Tabelle 17 listet die im Text verwendeten sprachlichen Merkmale auf, mit denen die Menschenähnlichkeit des Cobots variiert wurde.

Tabelle 17: Sprachliche Merkmale zum Framing der Menschenähnlichkeit des Cobots.

	Menschenähnliches Framing	Maschinelles Framing
Name (Darling et al., 2015)	<i>Paul</i>	<i>UR-5</i>
Zugeschriebene Eigenschaften (L. J. Baker et al., 2018)	<ul style="list-style-type: none"> • Paul braucht nur einen <i>Ru-hetag</i> pro Jahr, an dem er <i>fit gemacht</i> wird. • <i>Sprechen</i> kann Paul nicht. • Wenn die Mitarbeiter eine Frühstückspause einlegen, kommt Paul zwar nicht mit, wartet aber <i>artig</i> [...] im Stillstand. 	<ul style="list-style-type: none"> • Der UR-5 muss nur einen Tag pro Jahr gewartet werden. • Der UR-5 verfügt nicht über eine Sprachausgabe. • Wenn die Mitarbeiter eine Frühstückspause einlegen, verbleibt der UR-5 [...] im Stillstandsmodus.
Zugehörigkeit/ Gemeinschaft (Kory Westlund et al., 2016; Reeves & Nass, 1998, S. 156f.)	<ul style="list-style-type: none"> • Sie werden [...] gemeinsam mit dem kollaborierenden Roboter Paul zusammenarbeiten. 	<ul style="list-style-type: none"> • Sie werden [...] mithilfe des kollaborierenden Roboters UR-5 arbeiten.
Grad der Autonomie	Akteur, aktive Sprache:	Objekt, passive Sprache:

	Menschenähnliches Framing	Maschinelles Framing
(Stenzel, Chinellato, Bou et al., 2012)	<ul style="list-style-type: none"> • Paul würde [...] [Zusammenstöße] auch sofort bemerken und bliebe dann stehen. • Paul [ist] lernfähig und entwickelt sich weiter. 	<ul style="list-style-type: none"> • Die Sensoren des Roboters würden [...] [Zusammenstöße] auch sofort detektieren und den UR-5 stoppen. • Die programmierte Logik [ist] selbstlernend und entwickelt sich weiter.
Personalpronomen (Coeckelbergh, 2010b; Kory Westlund et al., 2016)	<i>er</i>	<i>der Roboter</i>

Analog dazu stellt Tabelle 18 diejenigen sprachlichen Merkmale dar, die die Mensch-Cobot-Relation unterschiedlich umschrieben.

Tabelle 18: Sprachliche Merkmale zum Framing der Kooperationsorientierung der Mensch-Cobot-Relation.

	Kooperierende Relation	Konkurrierende Relation
Hierarchie	<p>Cobot übernimmt Aufgaben, bei denen er für den Menschen nützlich ist.</p> <ul style="list-style-type: none"> • Bei Ihrer zukünftigen Arbeit wird Sie [der Roboter] bei denjenigen Teilen Ihrer Tätigkeiten unterstützen, bei welchen [...] [er] besonders nützlich ist. 	<p>Cobot übernimmt Aufgaben, bei denen Menschen über minderwertige Fähigkeiten verfügen.</p> <ul style="list-style-type: none"> • Bei Ihrer zukünftigen Arbeit wird [der Roboter] diejenigen Teile Ihrer Aufgaben übernehmen, bei denen Sie [...] [dem Roboter] unterlegen sind.
Vorzüge	<p>Vorteile von gemeinsamen Mensch-Roboter-Teams werden hervorgehoben.</p> <ul style="list-style-type: none"> • Die sinnvolle Kombination der Stärken von Mensch und Roboter bringt deutliche Vorteile. 	<p>Vorteile des Robotereinsatzes werden hervorgehoben.</p> <ul style="list-style-type: none"> • Die sinnvolle Nutzung der Stärken des Roboters bringt deutliche Vorteile.
Auswirkung auf Arbeitsweise	<p>Mitarbeitende werden durch den Cobot <i>motiviert</i>, ihre Arbeitsschwindigkeit so anzupassen,</p>	<p>Mitarbeitende fühlten sich durch die hohe Arbeitsschwindigkeit des Cobots <i>unter Druck gesetzt</i> und unternahmen</p>

	Kooperierende Relation	Konkurrierende Relation
	dass sie auf Augenhöhe arbeiten können.	große Anstrengungen, um mithalten zu können.
Mehrwert für Unternehmen	<p>Wert der Mitarbeitenden für das Unternehmen wurde unterstrichen.</p> <ul style="list-style-type: none"> • Der Mensch ist – im Zusammenspiel mit dem Roboter – für ein zukunftsorientiertes Unternehmen unverzichtbar. 	<p>Wert des Cobots für das Unternehmen wurde unterstrichen.</p> <ul style="list-style-type: none"> • Der Roboter ist – im Zusammenspiel mit dem Menschen – für ein zukunftsorientiertes Unternehmen unverzichtbar.

Die Texte hatten jeweils eine vergleichbare Länge und differierten nur leicht in der Lesbarkeit, die mittels des Flesch-Lesbarkeitsindex⁷ ermittelt wurde. Diese Unterschiede waren durch das Framing bedingt, bspw. durch sprachlich umständlichere Passivkonstruktionen im Vergleich zu Aktivformulierungen. Tabelle 19 stellt die Länge und Lesbarkeit der jeweiligen Texte gegenüber. Die vollständigen Texte im Wortlaut finden sich in Anhang C.I und C.II.

Tabelle 19: Länge und Lesbarkeit der verwendeten Framing-Texte in den jeweiligen Versuchsbedingungen (Experimentalstudie 1).

	Wahrgenommene Menschenähnlichkeit			Wahrgenommene Kooperationsorientierung		
	Menschen-ähnlich	Maschinell	Kontrollbed.	Kooperierend	Konkurrierend	Kontrollbed.
Flesch-Wert	56	47	27	39	37	32
Lesbarkeit	anspruchsvoll	schwierig	schwer	schwierig	schwierig	schwierig
Sätze	18	18	16	8	8	10
Wörter	254	264	264	143	161	135
Silben	479	536	616	302	338	311
Zeichen	1733	1878	2050	1040	1171	1037

⁷ Die Lesbarkeitswerte wurden online unter <https://fleschindex.de/formel> ermittelt.

4.5.2.3 Messmethodik

Zur Messung des Vertrauens und Misstrauens in den Cobot als AV wurde auf die TAS nach Jian et al. (2000) zurückgegriffen, die in diesem Forschungsgebiet ein gängiges Multi-Item-Messinstrument darstellt. Für diese Skala liegt eine validierte deutsche Fassung nach Pöhler et al. (2016) vor (vgl. Kapitel 3.4.9 zur Vertrauensmessung). Im Gegensatz zum Original umfasst diese deutsche Version nur elf anstatt zwölf Items, da das Item *Ich bin sicher im Umgang mit dem System* aufgrund einer geringen Trennschärfe, Faktorladung und Kommunalität aus der Gesamtskala entfernt wurde (Pöhler et al., 2016, S. 158). Die Bewertung der Items erfolgt anhand einer siebenstufigen Likert-Skala von 1 = *trifft gar nicht zu* bis 7 = *trifft völlig zu*. Die sechs positiv formulierten Items wurden durch Mittelwertbildung zu einem Wert für das Vertrauen und die fünf negativ formulierten zu einem Wert für das Misstrauen zusammengefasst.

Tabelle 20: Items der deutschen Fassung der TAS nach Ersetzung des Begriffs *System* durch *Roboter*.

Item	Konstrukt
Der Roboter arbeitet tadellos.	Vertrauen
Der Roboter bietet Sicherheit.	Vertrauen
Der Roboter ist verlässlich.	Vertrauen
Der Roboter ist vertrauenswürdig	Vertrauen
Ich kenne mich mit dem Roboter aus.	Vertrauen
Ich kann dem Roboter vertrauen.	Vertrauen
Der Roboter ist irreführend.	Misstrauen
Der Roboter verhält sich undurchsichtig.	Misstrauen
Die Handlungen des Roboters haben negative Auswirkungen zur Folge.	Misstrauen
Ich muss vorsichtig im Umgang mit dem Roboter sein.	Misstrauen
Ich misstrau den Entscheidungen des Roboters.	Misstrauen

Im Original misst die TAS das Vertrauen in ein nicht näher spezifiziertes System. Da die Studie die Messung des Vertrauens und Misstrauens in einen konkreten Cobot avisierte, wurde dieser allgemeine Begriff in den präsentierten Items vor dem Framing durch *Roboter* (vgl. Tabelle 20) bzw. nach dem Framing durch die konkrete Roboterbezeichnung je nach Versuchsbedingung (*UR-5/Paul*) ersetzt. Lediglich für die Teilnehmenden in der Kontrollgruppe blieb die Bezeichnung *Roboter* während der gesamten Befragung bestehen. Tabelle 21 illustriert die Unterschiede in den Formulierungen vor bzw. nach dem Framing bei der Darbietung der TAS. Die einzelnen Items innerhalb der Fragenbatterie wurden den Versuchsteilnehmenden in randomisierter Reihenfolge dargeboten, um Verzerrungen durch Reihenfolgeeffekte vorzubeugen.

Tabelle 21: Textfragmente zur Abfrage des Vertrauens und Misstrauens mittels TAS jeweils vor und nach dem Framing.

	Vor dem Framing	Nach dem Framing
Einführung Cobot	Kollaborierende Roboter sind in der Regel relativ kleine Roboter, die durch integrierte Sicherheitsmerkmale die direkte Zusammenarbeit zwischen Mensch und Roboter (quasi "Hand in Hand") ermöglichen.	– <i>nicht mehr notwendig</i> –
Bezug zu konkretem Roboter	Die folgenden Fragen beziehen sich auf einen konkreten kollaborierenden Roboter. Zu diesem werden Sie später weitere Informationen erhalten. Wie stellen Sie sich eine Zusammenarbeit mit einem solchen Roboter vor? Bitte markieren Sie bei jeder Aussage diejenige Ausprägung, welche Ihrer persönlichen Einschätzung am ehesten entspricht.	Stellen Sie sich vor, der <i>UR-5/Paul</i> [je nach Bedingung] wird in Ihrem Unternehmen eingesetzt. Bitte bewerten Sie auf Basis der zuvor erhaltenen Informationen.
Anweisung zur Beantwortung der Fragen	Es gibt keine richtige oder falsche Antwort. Antworten Sie bitte spontan nach Ihrem ersten Eindruck.	Es gibt keine richtige oder falsche Antwort. Antworten Sie bitte spontan nach Ihrem ersten Eindruck.
Beispielhaftes Item	Ich kenne mich mit dem Roboter aus.	Ich kenne mich mit dem <i>UR-5/Paul</i> [je nach Bedingung] aus.

Zur Erhebung der wahrgenommenen Menschenähnlichkeit als UV mittels Selbstauskünften wurde eine eigene Übersetzung der Subskala zur Menschlichkeit aus der überarbeiteten *Godspeed*-Skala von Ho und MacDorman (2017) verwendet. Die ursprüngliche Version stammt von Bartneck, Kulić, Croft und Zoghbi (2009) und stellt eine gängige Erhebungsmethode für die Wahrnehmung eines Roboters dar (Craenen, Deshmukh, Foster & Vinciarelli, 2018, S. 626), die neben der verwendeten Skala zur Menschlichkeit (*humanness*) weitere Subskalen zur Unheimlichkeit (*eeriness*), zum Gruselfaktor (*spine-tingling*) und zur Attraktivität (*attractiveness*) eines Roboters umfasst (Ho & MacDorman, 2017). Die verwendete Fragenbatterie zur Menschlichkeit verfügt über eine hohe interne Reliabilität

(Cronbachs $\alpha = .87$ aus Ho & MacDorman, 2017, S. 136). Sie umfasst fünf Items, die mittels eines siebenstufigen semantischen Differenzials von -3 bis +3 gemessen werden. Die englischsprachigen Items aus der Originalskala sowie die verwendeten deutschen Übersetzungen sind in Abbildung 4.8 dargestellt.

Wie stellen Sie sich [den Roboter] vor?		-3	-2	-1	0	1	2	3	
unbelebt <i>inanimate</i>		<input type="radio"/>	lebendig <i>living</i>						
künstlich <i>synthetic</i>		<input type="radio"/>	real <i>real</i>						
mechanische Bewegungen <i>mechanical movement</i>		<input type="radio"/>	Biologische Bewegungen <i>biological movement</i>						
menschengemacht <i>human-made</i>		<input type="radio"/>	menschenähnlich <i>human-like</i>						
ohne definierte Lebenszeit <i>without definite lifespan</i>		<input type="radio"/>	sterblich <i>mortal</i>						

Abbildung 4.8: Übersetzte Subskala der überarbeiteten *Godspeed*-Skala nach Ho und MacDorman (2017) als semantisches Differenzial zur Erhebung der wahrgenommenen Menschenähnlichkeit des Roboters.

Zur Steigerung der Reliabilität kam neben der Selbstauskunftsskala eine indirekte Messmethode zum Einsatz. Die Basis hierfür bildeten die Freitext-Beschreibungen des Cobots (mind. 40 Zeichen) durch die Versuchsteilnehmenden, die im Nachgang inhaltlich analysiert wurden. Zur quantitativen und intersubjektiven Durchdringung dieses sprachlichen Materials wurde ein Kodierleitfaden mit präzisen Handlungsanweisungen und Kriterien entwickelt, wie es bspw. aus der evaluativen qualitativen Inhaltsanalyse nach Mayring (2015) bekannt ist. Anhand dieses Leitfadens analysierten zwei unabhängige Gutachter (*rater*) die Aussagen der Teilnehmenden. Im Unterschied zum Vorgehen bei der inhaltlich-strukturierenden Inhaltsanalyse wurden nicht einzelne Textpassagen kodiert, sondern aufgrund der kurzen Beschreibungen eine Gesamtbewertung pro Fall vorgenommen (vgl. Schreier, 2014, S. 10f.). Die ersten Versionen des Kodierleitfadens wurden anhand eines kleinen Ausschnitts der Kommentare getestet und im Falle divergierender Bewertungen entsprechend präzisiert. Die Bewertungen erfolgten auf einer fünfstufigen Skala von $1 = \text{maschinell}$ bis $5 = \text{menschenähnlich}$. Die *inter-rater*-Reliabilität wurde mittels Krippendorffs α (Krippendorff, 1970) unter Verwendung des webbasierten Tools *ReCal OIR*

von Freelon (2013) ermittelt. Basierend auf der Bewertung mithilfe des Kodierleitfadens ergab sich letztlich ein ausreichend hoher Wert für Krippendorffs $\alpha = .824$. Verbleibende strittige Fälle zwischen den beiden Gutachtern wurden gemeinsam besprochen, um so schlussendlich eine einstimmige Einigung zu erzielen. Die beiden verwendeten Kodierleitfäden finden sich in Anhang C.III und C.IV.

Mangels bestehender validierter Instrumente wurde zur Bestimmung der wahrgenommenen Kooperationsorientierung als weitere UV eine eigene Skala entwickelt, die ebenfalls als semantisches Differenzial mit fünf Items umgesetzt wurde. Die verwendete Skala ist in Abbildung 4.9 dargestellt.

Mitarbeitende in der Produktion werden durch die Einführung [von Robotern] an ihrem Arbeitsplatz...								
	-3	-2	-1	0	1	2	3	
...belastet	<input type="radio"/>	...entlastet						
...mehr gestresst	<input type="radio"/>	...weniger gestresst						
...bedroht	<input type="radio"/>	...unterstützt						
...ersetzt	<input type="radio"/>	...ergänzt						

Abbildung 4.9: Eigens entwickeltes semantisches Differenzial zur Erhebung der wahrgenommenen Kooperationsorientierung.

Zur explorativen Analyse wurden ferner einige klassische Single-Item-Messmethoden aus den Technologieakzeptanzmodellen (vgl. Kapitel 3.4.8) integriert, die in dem gegebenen Kontext besonders interessant erschienen. Die verwendeten Items, die in Tabelle 22 aufgelistet sind, wurden auf einer fünfstufigen Zustimmungsskala von 1 = *stimme gar nicht zu* bis 5 = *stimme völlig zu* bewertet.

Tabelle 22: Verwendete Items zur Messung verschiedener Facetten der Akzeptanz (Experimentalstudie 1).

Konstrukt	Item	Quelle
Wahrgenommene Benutzerfreundlichkeit	Die Interaktion mit [Robotername] stelle ich mir einfach vor.	TAM
Wahrgenommener Nutzen	[Robotername] wäre für meine Arbeit nützlich.	TAM

Konstrukt	Item	Quelle
Nutzungsintention	Wenn ich wählen könnte, ob [Robotername] mich bei meiner Arbeit unterstützen soll, würde ich mich für die Unterstützung durch [Robotername] entscheiden.	TAM
Wahrgenommene Sicherheit	Bei der Nutzung von [Robotername] würde ich mich sicher fühlen.	TAM-HRC nach Bröhl et al. (2017)
Tatsächlicher Gebrauch	Ich würde [Robotername] gegenüber anderen Maschinen im industriellen Umfeld bevorzugen.	TAM
Angst vor Jobverlust	Ich würde befürchten, dass ich durch [Robotername] langfristig meinen Job verlieren werde.	TAM-HRC nach Bröhl et al. (2017) als Konstrukt <i>ethische Implikation</i>
Einstellung zur Nutzung I	Die Verwendung des [Robotername] ist eine gute Idee.	UTAUT
Einstellung zur Nutzung II	Mir würde es gefallen, mit dem [Robotername] zu arbeiten.	UTAUT
Willkommener Arbeitskollege	Ich hätte den [Robotername] gerne als Arbeitskollegen.	angelehnt an Weiss et al. (2008)

Ferner wurde die initiale Einstellung der Teilnehmenden gegenüber Robotern mittels der *negative attitude towards robots scale* (NARS) nach Nomura, Kanda, Suzuki und Kato (2004) abgefragt. Hierbei handelt es sich um eine validierte und gängige Messmethode im MRI-Bereich (Krägeloh, Bharatharaj, Sasthan Kutty, Nirmala & Huang, 2019, S. 90; Weiss et al., 2016, S. 8). Die 14 Items der Skala erheben die negative Einstellung in Bezug auf Interaktionssituationen mit Robotern (sechs Items), den sozialen Einfluss durch Roboter (fünf Items) und Emotionen bei der MRI (drei Items). In der vorliegenden Studie wurde die von Bartneck (2019) online zur Verfügung gestellte deutsche Übersetzung verwendet, die mehrere Zyklen der Rückübersetzung und Überarbeitung durchlaufen hat (Bartneck, Suzuki, Kanda & Nomura, 2006, S. 221). Die Bewertung der Items erfolgte anhand einer fünfstufigen Zustimmungsskala, die insbesondere für solche Batterien an Statements in der *Ich*-Form am vielseitigsten einsetzbar ist (Rohrman, 1978, S. 240). Die verbalen Verankerungen der Skalenwerte entstammen der von Rohrman (1978) empirisch entwickelten Intensitätsskala, wobei die mittlere Antwortkategorie *mittelmäßig* in Anlehnung an Hofmann, Möckelmann und Walach (2003, S. 5) durch *stimme teils zu, teils nicht* ersetzt wurde.

Ferner wurde die Wahrnehmung der eigenen Kompetenz im Umgang mit neuen Technologien auf einer fünfstufigen Zustimmungsskala mit den Ausprägungen *stimme gar nicht / wenig / teils / überwiegend / völlig zu* gemessen und bei der Auswertung als quasi-metrisch interpretiert. Außerdem wurden die Versuchsteilnehmenden dazu befragt, wie intensiv sie sich bisher mit Robotern auseinandergesetzt haben, beispielsweise durch die Rezeption von Filmen, Zeitungsartikeln oder Internetquellen. Die Einschätzung erfolgte auf einer fünfstufigen Intensitätsskala nach Rohrmann (1978, S. 231) mit folgenden verbalen Verankerungen: 1 = *nicht intensiv*, 2 = *wenig intensiv*, 3 = *mittelmäßig intensiv*, 4 = *ziemlich intensiv*, 5 = *sehr intensiv*.

Die interne Konsistenz sämtlicher Skalen wurde anhand Cronbachs α gemessen. Mit Ausnahme der *Godspeed*-Skala ($\alpha > .70$) erreichen alle Skalen eine ausreichende ($\alpha > .80$), im Falle der Akzeptanz-Items sogar eine hohe Reliabilität ($\alpha > .90$) (Döring & Bortz, 2016, S. 443). Der geringere Wert für die *Godspeed*-Skala lässt sich messtheoretisch auf die vergleichsweise geringe Anzahl an Items zurückführen. Generell ist darauf hinzuweisen, dass die Bestimmung der internen Konsistenz auf Basis einer einzelnen Testung u. U. zu einer Überschätzung der Reliabilität führen kann und sich eine rein zahlenbasierte, kontextfreie Interpretation verbietet (Döring & Bortz, 2016, S. 468f.). Die gemessenen Reliabilitätswerte sämtlicher verwendeter Multi-Item-Skalen sind in Anhang C.VIII.a aufgelistet.

4.5.2.4 Studienablauf

Die Versuchsteilnehmenden wurden über den Panel-Anbieter *ToLuna Deutschland GmbH* auf die Startseite des Online-Experiments geleitet. Tabelle 23 verdeutlicht den Ablauf des Experiments.

Tabelle 23: Tabellarische Darstellung des Ablaufs von Experimentalstudie 1.

Nr.	Schritt	Beschreibung
1	Einführung zur Studie	Kurzinformation über die Studie.
2	Screening-Fragen	Vier Multiple Choice-Fragen zur Überprüfung der Zugehörigkeit der Teilnehmenden zur Zielpopulation.
3	NARS	Präsentation der Items in randomisierter Reihenfolge.
4	Thematische Einführung zum Thema Cobots	Kurze fachliche Einführung zu Cobots. Außerdem Information zum Gegenstand der Studie, insb. mit dem Hinweis darauf, dass sich die Studie auf einen konkreten Cobot bezieht, den sich die

Nr.	Schritt	Beschreibung
		Versuchsteilnehmenden vorstellen sollen und der im Verlauf noch näher beschrieben wird.
5	Prä-Framing Vertrauensmessung (TAS)	Die Items bezogen sich auf einen konkreten Cobot und wurden in randomisierter Reihenfolge präsentiert.
6	Zusätzliche Fragen (Demografie, relevante Grundeinstellungen)	Geschlecht, Alter, Berufserfahrung, Branche des Unternehmens, Kompetenzüberzeugung im Umgang mit neuen Technologien sowie die Intensität der bisherigen Auseinandersetzung mit Robotern (z. B. in Filmen, Zeitungsartikeln oder in Internetquellen).
7	Randomisierte Gruppenzuweisung	<i>Siehe Versuchsgruppen in Kapitel 4.5.1.</i>
8	Präsentation der Framing-Texte	Die Teilnehmenden wurden explizit aufgefordert, die Texte gründlich zu lesen und konnten frühestens nach 20 Sekunden Verweildauer pro Text auf die nächste Seite navigieren.
9	Verständnisfragen	Abfrage des im Text genannten Namens sowie der Vorteile des Cobots in zwei Multiple-Choice-Fragen mit je vier bzw. fünf Antwortmöglichkeiten. Bei mindestens einer falschen Antwort erfolgte ein Ausschluss aus der Studie.
10	Manipulationskontrolle	<i>Siehe verwendete Messmethodik zur Manipulationskontrolle in Kapitel 4.5.2.3.</i>
11	Post-Framing Vertrauensmessung (TAS)	Die Items bezogen sich auf einen konkreten Cobot und wurden in randomisierter Reihenfolge präsentiert.
12	Abfrage von Akzeptanz-Items	Einschätzung ausgewählter Akzeptanz-Items.
13	Freitext-Beschreibung des Cobots	Mindestens 40 Zeichen.
14	Generelles Feedback als Freitext	Optional.
15	Dank für Teilnahme / Absenden der Daten / Weiterleitung zum Panel-Anbieter	Automatische Entlohnung durch Panel-Anbieter.

Mit Ausnahme des generellen Feedbacks am Ende handelte es sich bei sämtlichen Fragen um Pflichtfragen. Die Versuchsteilnehmenden benötigten im Durchschnitt $M = 12:51$ Minuten ($SD = 6:51$ Minuten; $Mdn = 11:07$ Minuten) für die komplette Bearbeitung der Studie, bezogen auf alle gültigen Datensätze.

4.5.2.5 Statistische Analysen

Grundsätzlich stehen zur Analyse von Prä-Post-Studiendesigns wie in der vorliegenden Experimentalstudie die folgenden statistischen Auswertungsmethoden zur Verfügung (Knapp & Schafer, 2009):

1. gemischte Varianzanalyse (*mixed ANOVA*) mit Zwischensubjektfaktoren und Messwiederholungsfaktor,
2. Zugewinnanalyse (*gain score analysis*) als *between subject-ANOVA* mit der Differenz aus Vertrauenswert nach und vor dem Framing als AV und entsprechend ohne Einbezug eines Messwiederholungsfaktors,
3. Kovarianzanalyse (*analysis of covariance*; ANCOVA) mit Zwischensubjektfaktoren, dem Post-Wert als AV und dem Prä-Wert als Kovariate.

Die jeweiligen Verfahren verfügen über verschiedene Stärken und Schwächen. Folglich werden deren geeignete Anwendungsgebiete fortlaufend diskutiert. Die gemischte ANOVA und die weniger komplexe Zugewinnanalyse sind letztlich mathematisch äquivalent (Knapp & Schafer, 2009). Die Zugewinnanalyse und die ANCOVA beantworten allerdings leicht unterschiedliche Forschungsfragen und können kontraintuitiv u. U. zu verschiedenen statistischen Ergebnissen führen. Dieses Phänomen wird in der Literatur unter dem Begriff des *Lord-Paradoxons* (Lord, 1967) diskutiert (u. a. in Wainer, 1991).

In der vorliegenden Studie wird zunächst eine gemischte $2 \times 2 \times 2$ -ANOVA berechnet, um den grundsätzlichen Einfluss der experimentellen Manipulation nachzuweisen. Zur genaueren Exploration der Einflüsse der Zwischensubjektfaktoren auf die AVs wird im weiteren Verlauf die besser interpretierbare Zugewinnanalyse eingesetzt. Da die *Veränderung* der Vertrauenswerte *durch die Intervention* anstatt einer Interpretation der *absoluten* Vertrauenswerte je Versuchsgruppe nach dem Framing im Fokus des Interesses steht, ist dieses Verfahren im Vergleich zur ANCOVA besser geeignet. Außerdem ist durch die randomisierte Zuweisung zur Versuchsgruppe eine (zumindest annähernde) Gleichverteilung der Prä-Vertrauenswerte anzunehmen. Daraus ergibt sich ein vollfaktorielles 2×2 -*between subject*-Studiendesign mit zwei je zweistufigen Zwischensubjektfaktoren, nämlich der wahrgenommenen Menschenähnlichkeit des Cobots und der wahrgenommenen Kooperationsorientierung der Mensch-Cobot-Relation. Als quasi-metrische AV wird jeweils die Vertrauensveränderung bzw. die Misstrauensveränderung herangezogen, die als Differenz zwischen den gemessenen Vertrauens- und Misstrauenswerten zu den beiden Messzeitpunkten je vor und nach dem Framing bestimmt wird.

Da sich nach einer ersten Analyse zeigte, dass das Vertrauen in allen vier Experimentalgruppen zum zweiten Messpunkt höher lag, wurde eine Kontrollgruppe nacherhoben, um auszuschließen, dass sich dieser Anstieg auf einen *test-retest bias* zurückführen lässt. Die Vertrauensentwicklungen in der Kontrollgruppe wurden mithilfe von *t*-Tests mit denjenigen in den Experimentalgruppen verglichen. Bei der Manipulationskontrolle kamen ebenfalls *t*-Tests sowie Mann-Whitney-*U*-Tests für Gruppenvergleiche zum Einsatz. Experimentelle randomisierte Kontrollgruppenstudien mit Vorher-Nachher-Messungen der AV und aktiver experimenteller Manipulation gelten als Goldstandard in Hinblick auf die interne Validität und erlauben vergleichsweise eindeutige Nachweise von Kausalzusammenhängen (Döring & Bortz, 2016, S. 94).

Gemäß der Konvention wurde das Signifikanzniveau für die Signifikanztests *a priori* auf $\alpha = 5\%$ festgelegt (Döring & Bortz, 2016, S. 666). Da neuere Literatur einseitiges Testen zunehmend problematisiert (Lombardi & Hurlbert, 2009; Ruxton & Neuhäuser, 2010), wurden grundsätzlich zweiseitige Signifikanztests verwendet. Die Interpretation der Effektstärken und Korrelationskoeffizienten erfolgte anhand der Daumenregeln nach Cohen (1988).

Vor der Anwendung von parametrischen Auswertungsverfahren (wie z. B. unabhängiger *t*-Test, ANOVA, o. Ä.) wurden deren Voraussetzungen, nämlich Normalverteilung der AV in jeder Versuchsgruppe (Shapiro-Wilk-Test $p > .05$) sowie Varianzhomogenität (Levene-Test $p > .05$), getestet und können – sofern nicht abweichend angegeben – als gegeben angesehen werden. Falls keine Varianzhomogenität vorlag, kam Welchs *t*-Test zur Anwendung. Falls keine Normalverteilung vorlag, wurde der Mann-Whitney-*U*-Test als nichtparametrische Alternative verwendet.

Im Folgenden wird bei Mann-Whitney-*U*-Tests aufgrund der ausreichend hohen Stichprobenanzahl jeweils die asymptotische Signifikanz berichtet. Die Fehlerbalken in den Diagrammen repräsentieren jeweils den Standardfehler. Signifikante Unterschiede oder Zusammenhänge werden in Tabellen und Abbildungen durch Asterisken gekennzeichnet, wobei folgende Zuordnung gilt: * $p < .05$, ** $p < .01$, *** $p < .001$. Bei der Korrelationsanalyse wird jeweils die Pearson-Korrelation ausgewiesen und aufgrund des explorativen Charakters dieser Auswertungen auf eine Korrektur der Signifikanzwerte aufgrund des multiplen Testens verzichtet.

4.5.3 Ergebnisse

4.5.3.1 Manipulationskontrolle

Im Rahmen der Manipulationskontrolle wurde überprüft, ob die experimentelle Manipulation, d. h. die Präsentation der unterschiedlichen Framing-Texte, zu Unterschieden in der

wahrgenommenen Menschenähnlichkeit des Cobots und der wahrgenommenen Kooperationsorientierung der Mensch-Cobot-Relation über die verschiedenen Versuchsgruppen hinweg führte.

Ein zweiseitiger unabhängiger t -Test verdeutlichte signifikante Unterschiede zwischen den beiden Versuchsgruppen in Hinblick auf die wahrgenommene Menschenähnlichkeit des Cobots, $t(116) = 2.50$, $p = .014$, $\eta^2 = .051$. Gemäß der Einstufung nach Cohen (1988, S. 285f.) handelt es sich um einen kleinen bis mittelstarken Effekt. Wie Abbildung 4.10 zeigt, lag der Mittelwert für die empfundene Menschenähnlichkeit in der Versuchsgruppe, die das menschenähnliche Framing erhielt, mit $M = -0.45$, $SD = 1.23$, erwartungskonform signifikant höher als in der Versuchsgruppe, die das maschinelle Framing erhielt, $M = -1.02$, $SD = 1.22$. Gleichwohl weisen die negativen Mittelwerte darauf hin, dass der Cobot in beiden Versuchsgruppen eher maschinell anstatt menschenähnlich wahrgenommen wurde.

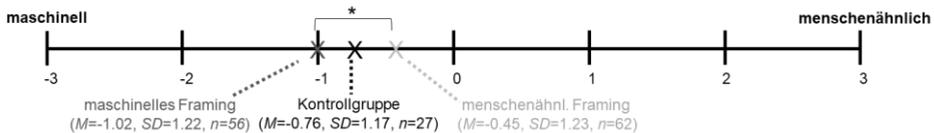


Abbildung 4.10: Grafische Darstellung der wahrgenommenen Menschenähnlichkeit (*Godspeed*-Skala) je Versuchsgruppe mit inferenzstatistischen Ergebnissen.

Zusätzlich zur Selbstauskunftsskala wurden die Freitext-Beschreibungen des Cobots durch die Versuchsteilnehmenden anhand eines Kodierleitfadens bewertet, um die vorigen Selbstauskunftswerte durch eine indirekte Messmethode abzusichern (vgl. Kapitel 4.5.2.3). Die Bewertungen erfolgten auf einer fünfstufigen Skala von $1 = \text{maschinell}$ bis $5 = \text{menschenähnlich}$. Ein durchgeführter Mann-Whitney- U -Test als non-parametrische Alternative zum unabhängigen t -Test zeigte signifikante Unterschiede zwischen den beiden Versuchsgruppen, $U = 1124.50$, $Z = -3.209$, $p = .001$. Die Effektstärke liegt mit $\eta^2 = .092$ im Bereich eines mittelstarken Effekts. Der Median für beide Gruppen lag jeweils bei $M = 3.00$. Die Mittelwerte sind in Abbildung 4.11 ersichtlich. Aus methodischer Sicht ist zu berücksichtigen, dass zwischen der Beantwortung der *Godspeed*-Skala und den Freitext-Beschreibungen die wiederholte Messung des Vertrauensniveaus stattfand. Diese eingeschobene Messung könnte zu Reihenfolgeeffekten auf die beiden Messungen des Vermenschlichungsniveaus geführt haben.

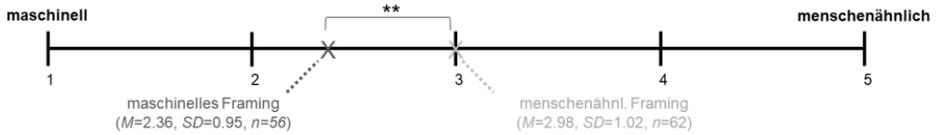


Abbildung 4.11: Grafische Darstellung der wahrgenommenen Menschenähnlichkeit (Bewertung gemäß Kodierleitfaden als indirekte Messmethode) je Versuchsgruppe mit inferenzstatistischen Ergebnissen.

In Hinblick auf die wahrgenommene Kooperationsorientierung zeigte ein Mann-Whitney- U -Test als non-parametrische Alternative zum unabhängigen t -Test signifikante Unterschiede zwischen den beiden Versuchsgruppen, $U = 1055.50$, $Z = -3.686$, $p < .001$. Wie in Abbildung 4.12 ersichtlich, repräsentierten die Versuchsteilnehmenden, die das konkurrierende Framing erhielten, die Mensch-Cobot-Relation mit $M = 0.36$, $SD = 1.56$ tatsächlich als konkurrenzorientierter als die Versuchsteilnehmenden, die ein kooperierendes Framing erhielten, $M = 1.43$, $SD = 1.08$. Es handelt sich hierbei mit $\eta^2 = .115$ um einen mittelstarken bis starken Effekt. In der Gruppe, die das konkurrierende Framing erhielt, lag auch der Median mit $Mdn = 0.50$ niedriger als in der Gruppe, die das kooperierende Framing erhielt, $Mdn = 1.50$.

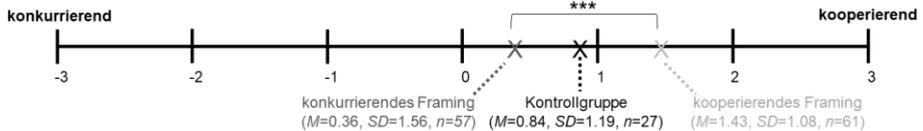


Abbildung 4.12: Grafische Darstellung der wahrgenommenen Kooperationsorientierung der Mensch-Cobot-Relation je Versuchsgruppe mit inferenzstatistischen Ergebnissen.

Sämtliche interferenzstatistischen Ergebnisse im Rahmen der Manipulationskontrolle legen nahe, dass die experimentelle Manipulation zu signifikanten Veränderungen in den UVs geführt hat. In Anhang V ist die Verteilung der Messwerte pro Versuchsgruppe basierend auf den Selbstauskunftsskalen zusätzlich als Box-Plot dargestellt.

4.5.3.2 Vertrauensentwicklung

Abbildung 4.13 visualisiert die Vertrauensentwicklung zwischen den beiden Messzeitpunkten pro Experimentalbedingung. Dabei zeigt sich in allen Bedingungen deskriptiv ein Vertrauenszuwachs, der mit Ausnahme der maschinell-kooperierenden Versuchsbedingung signifikant ausfällt.

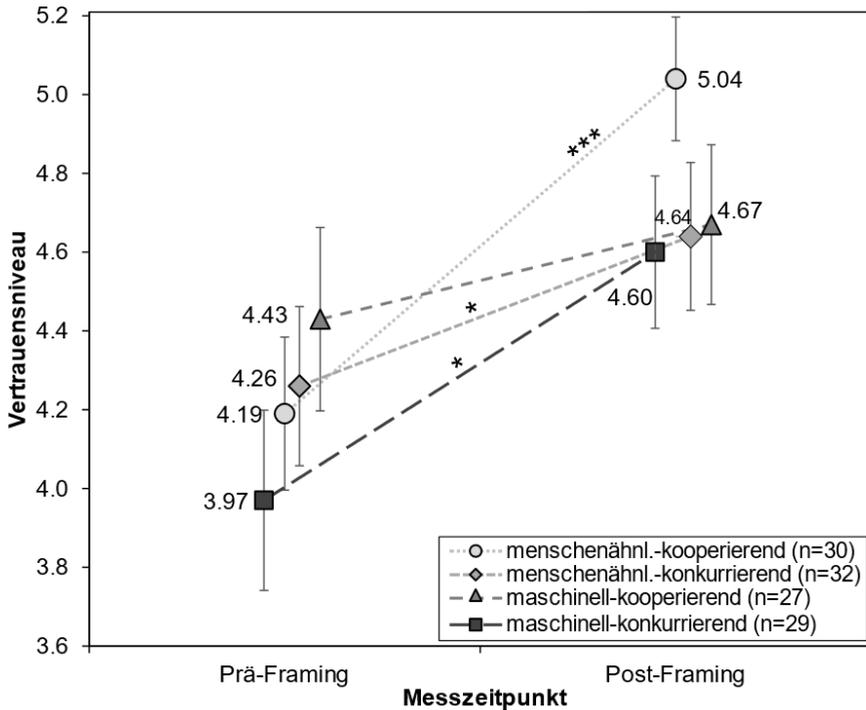


Abbildung 4.13: Grafische Darstellung des gemessenen Vertrauensniveaus an beiden Messzeitpunkten je Versuchsgruppe.

Die detaillierten inferenzstatistischen Ergebnisse zur Vertrauensentwicklung auf Basis je eines verbundenen *t*-Tests pro Versuchsbedingung finden sich in Tabelle 24.

Tabelle 24: Ergebnisse der inferenzstatistischen Analyse zur Veränderung des Vertrauensniveaus zwischen beiden Messzeitpunkten je Versuchsgruppe.

	<i>t</i> -Test		
	<i>df</i>	<i>t</i>	<i>p</i>
menschenähnlich-konkurrierend	29	-4.940	< .001***
menschenähnlich-konkurrierend	31	-2.613	.014*
maschinell-konkurrierend	26	-1.442	.161
maschinell-konkurrierend	28	-2.603	.015*

4.5.3.3 Misstrauensentwicklung

Abbildung 4.14 visualisiert die Misstrauensentwicklung zwischen den beiden Messzeitpunkten pro Experimentalbedingung. Dabei zeigt sich in allen Bedingungen eine signifikante Abnahme des Misstrauensniveaus zwischen den Messzeitpunkten.

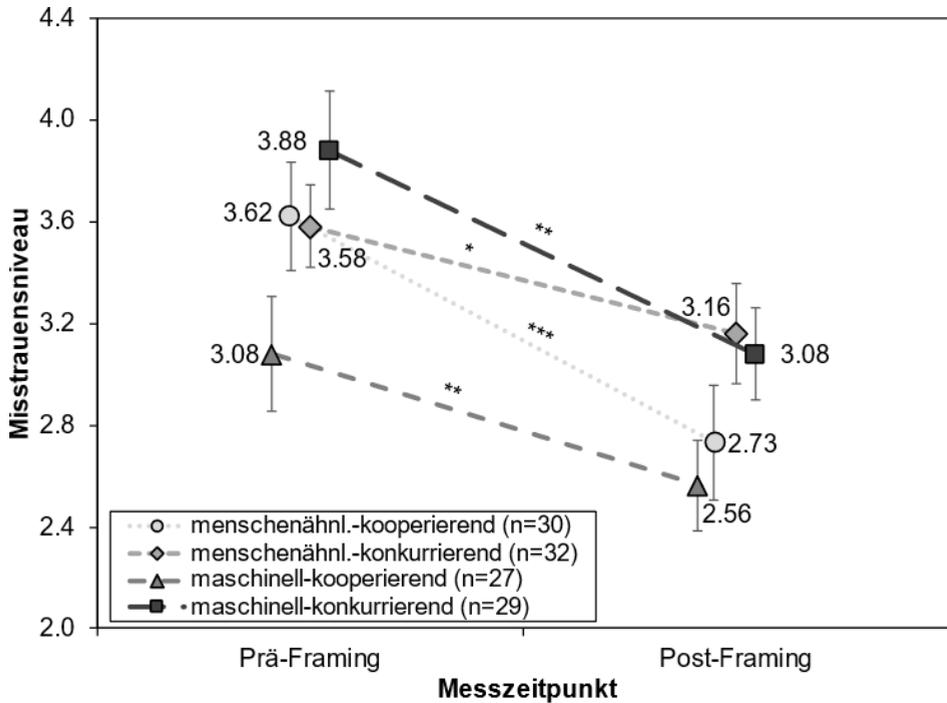


Abbildung 4.14: Grafische Darstellung des gemessenen Misstrauensniveaus an beiden Messzeitpunkten je Versuchsgruppe.

Die detaillierten inferenzstatistischen Ergebnisse zur Misstrauensentwicklung auf Basis je eines verbundenen *t*-Tests pro Versuchsbedingung finden sich in Tabelle 25.

Tabelle 25: Ergebnisse der inferenzstatistischen Analyse zur Veränderung des Misstrauensniveaus zwischen beiden Messzeitpunkten je Versuchsgruppe.

	<i>t</i> -Test		
	<i>df</i>	<i>t</i>	<i>p</i>
menschenähnlich-kooperierend	29	4.174	< .001***
menschenähnlich-konkurrierend	31	2.618	.014*
maschinell-kooperierend	26	3.095	.005**
maschinell-konkurrierend	28	3.361	.002**

4.5.3.4 Vergleich mit Kontrollgruppe

Der Vergleich zwischen Experimental- und Kontrollgruppe diente dazu, eine Verzerrung der gemessenen Vertrauens- und Misstrauensänderungen in den Experimentalgruppen durch einen *test-retest bias* ausschließen zu können. Abbildung 4.15 zeigt die Vertrauensentwicklung über die beiden Messzeitpunkte hinweg einerseits in der Kontrollgruppe ohne roboterbezogenes Framing ($n = 27$) und andererseits in den Experimentalgruppen ($n = 118$). Für diese Auswertung wurden die Versuchsteilnehmenden in den vier Experimentalgruppen per Mittelwertbildung zu einer kumulierten Experimentalgruppe zusammengefasst. Zwei verbundene *t*-Tests belegten, dass sich das Vertrauen nur in der Experimentalgruppe signifikant erhöhte, $t(117) = -5.68$, $p < .001$, wohingegen in der Kontrollgruppe keine signifikante Veränderung feststellbar war, $t(26) = -1.21$, $p = .234$.

Bei Betrachtung der Misstrauensentwicklung zeigte sich ein vergleichbares Muster, das sich auch in den signifikanten negativen Korrelationen zwischen Vertrauen und Misstrauen sowohl vor dem Framing, $r = -.508$, $p < .001$, als auch nach dem Framing, $r = -.535$, $p < .001$, widerspiegelt. In der Kontrollgruppe ließ sich sogar eine leichte, nicht-signifikante Zunahme des Misstrauens um $M = 0.10$ feststellen, $t(26) = -0.67$, $p = .507$, wohingegen das Misstrauen in den Experimentalgruppen signifikant abnahm, $t(117) = 6.62$, $p < .001$.

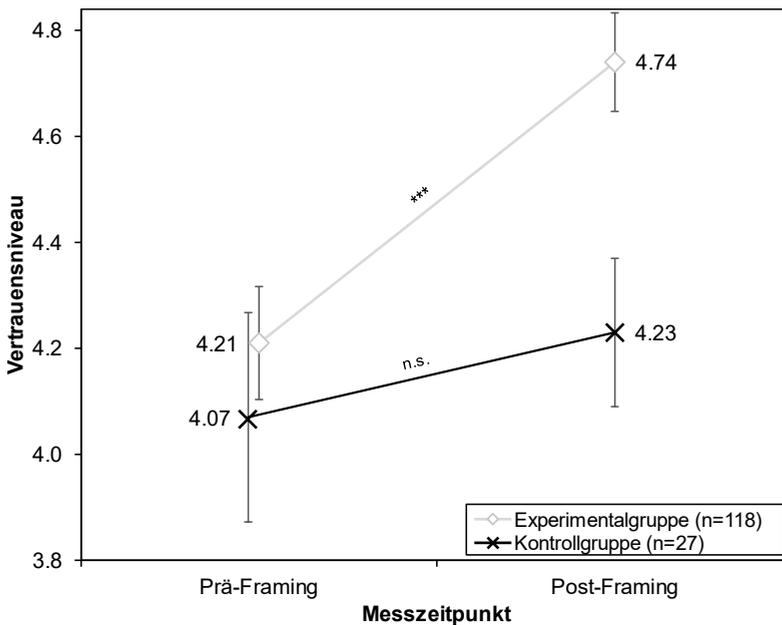


Abbildung 4.15: Entwicklung des Vertrauensniveaus zwischen beiden Messzeitpunkten in der Experimentalgruppe (alle Versuchsbedingungen, die eine experimentelle Manipulation erhielten, gemittelt) und der Kontrollgruppe.

Einschränkend ist hierbei zu erwähnen, dass die jeweilige Teststärke der durchgeführten t -Tests aufgrund der stark unterschiedlichen Gruppengröße variiert, was zu einer erhöhten Wahrscheinlichkeit eines Fehlers zweiter Art bei der Kontrollgruppe führt. Daher wurde ergänzend ein zweiseitiger Welch t -Test durchgeführt, um zu prüfen, ob sich die jeweiligen Vertrauensveränderungen in Experimental- und Kontrollgruppe voneinander unterscheiden. Der Test zeigte einen signifikanten Unterschied zwischen beiden Gruppen in Hinblick auf die Vertrauensveränderung über die Messzeitpunkte hinweg, $t(48.5) = -2.67, p = .010$.⁸

Diese Analyseergebnisse erlauben die Schlussfolgerung, dass die Vertrauenssteigerungen in den Experimentalgruppen zumindest zum überwiegenden Teil durch die experimentelle Intervention, d. h. das Lesen der cobot-bezogenen Texte, verursacht wurden. Ein möglicher *test-retest bias* schlägt sich allenfalls marginal nieder. Die Kontrollgruppe wird daher in den folgenden Analysen nicht weiter berücksichtigt. Somit liegt für die weiteren Analysen ein vollfaktorielles Design mit zwei je zweistufigen Zwischensubjektfaktoren vor.

4.5.3.5 Einfluss der UVs auf Vertrauen

Der Einfluss des Framings auf das Vertrauen in den Cobot wurde zunächst mittels einer mehrfaktoriellen gemischten $2 \times 2 \times 2$ -ANOVA untersucht (vgl. Kapitel 4.5.2.5). Der Messzeitpunkt mit den Ausprägungen vor dem Framing und nach dem Framing fungiert als zweistufiger Innersubjektfaktor. Die zwei jeweils zweistufigen Zwischensubjektfaktoren beziehen sich auf die UVs, d. h. die wahrgenommene Menschenähnlichkeit und die wahrgenommene Kooperationsorientierung. Es zeigt sich ein signifikanter Interaktionseffekt zwischen allen drei Faktoren: $F_{\text{Interaktion}}(1,114) = 5.40, p = .022, \eta^2 = .045$. Zusätzlich ergibt sich ein signifikanter Haupteffekt des Messzeitpunktes, $F_{\text{Messzeitpunkt}}(1,114) = 32.50, p < .001, \eta^2 = .222$, der allerdings aufgrund des vorhandenen Interaktionseffekts nur bedingt interpretierbar ist.

Zur weiteren Exploration des gefundenen Interaktionseffekts wurde eine 2×2 -Zwischensubjekt-ANOVA mit der Vertrauensveränderung als Differenz zwischen dem Vertrauensniveau nach dem Framing und vor dem Framing als AV durchgeführt. Erwartungsgemäß zeigte sich rechnerisch erneut der bereits zuvor identifizierte Interaktionseffekt zwischen beiden UVs, $F_{\text{Interaktion}}(1,114) = 5.40, p = .022, \eta^2 = .045$. Die Haupteffekte für beide UVs waren nicht signifikant, $F_{\text{Menschenähnlichkeit}}(1,114) = 0.99, p = .323$; $F_{\text{Kooperationsorientierung}}(1,114) = 0.045, p = .832$. Zur genaueren Exploration des Interaktionseffekts wurden *post hoc-t*-Tests durchgeführt, deren Ergebnisse in den Interaktionsdiagrammen in Abbildung 4.16 und in Abbildung 4.17 dargestellt sind. Grafisch lässt sich dabei erkennen, dass

⁸ Um eine Normalverteilung der Variablen zu erreichen, mussten weitere sieben Ausreißer aus dem Datensatz entfernt werden. Der Mann-Whitney- U -Test als non-parametrische Alternative ergab keinen signifikanten Unterschied, $U = 1949.00, Z = 1.811, p = .070, R^2 = .023$, weist allerdings bei einem angenommenen mittleren Effekt der Stärke $d = 0.5$ auch nur eine Teststärke von $(1-\beta) = 62.3\%$ auf.

es sich um eine disordinale Interaktion handelt, die keine globale Interpretation der Haupteffekte pro UV erlaubt (Döring & Bortz, 2016, S. 713).

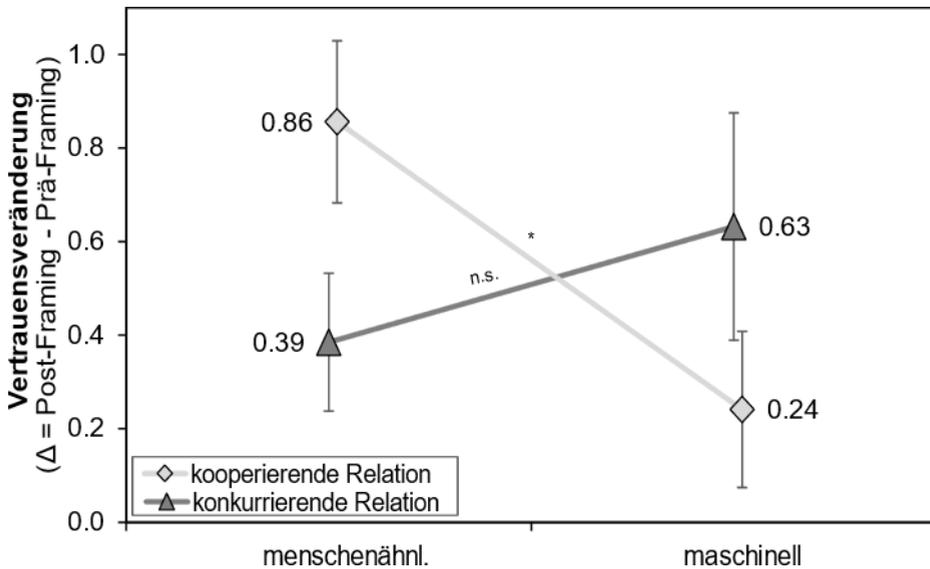


Abbildung 4.16: Interaktionsdiagramm zur Vertrauensveränderung zwischen den beiden Messzeitpunkten (mit wahrgenommener Menschenähnlichkeit abgetragen auf der horizontalen Achse).

In Abbildung 4.16 zeigt sich der einzige signifikante Unterschied, der sich im Rahmen der *post hoc*-*t*-Tests ergab. Im Falle einer kooperierenden Mensch-Cobot-Relation entwickelt sich das Vertrauen in einen menschenähnlich wahrgenommenen Cobot signifikant positiver als in einen maschinell wahrgenommenen Cobot, $t(114) = 0.16$, $p = .023$, $\eta^2 = .045$. Im Falle einer konkurrierenden Mensch-Cobot-Relation ergibt sich deskriptiv ein umgekehrtes Muster, d. h. die Vertrauensentwicklung fällt für einen maschinell wahrgenommenen Cobot positiver aus. Allerdings besteht hierbei kein signifikanter Unterschied in der Vertrauensentwicklung zwischen einem menschenähnlich und einem maschinell wahrgenommenen Cobot, $t(114) = -0.96$, $p = .340$, $\eta^2 = .008$. Wie im Interaktionsdiagramm in Abbildung 4.17 ersichtlich, liegen keine weiteren signifikanten Unterschiede auf Basis der *post hoc*-*t*-Tests vor.

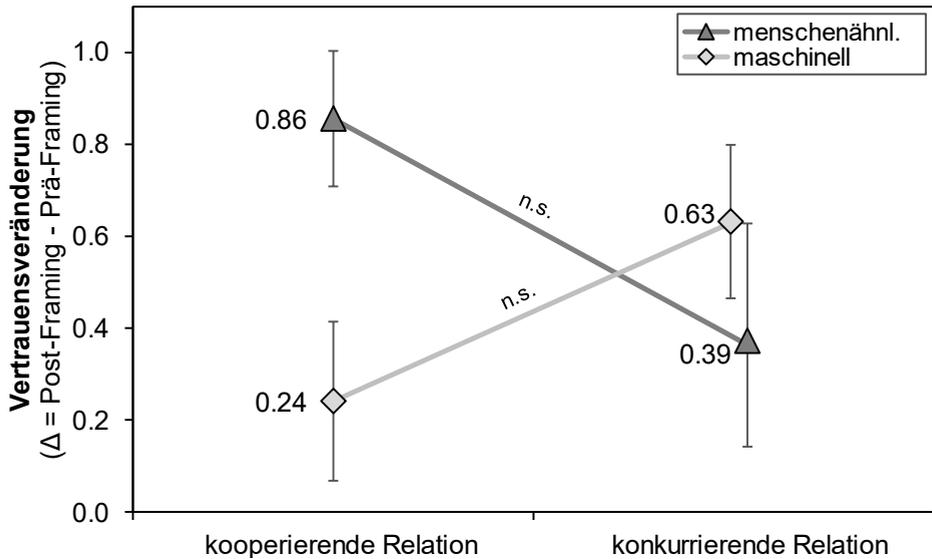


Abbildung 4.17: Interaktionsdiagramm zur Vertrauensveränderung zwischen den beiden Messzeitpunkten (mit wahrgenommener Kooperationsorientierung abgetragen auf der horizontalen Achse).

Abbildung 4.18 stellt die Vertrauensveränderungen pro Versuchsgruppe zwischen den beiden Messzeitpunkten dar. Erneut lässt sich der bereits thematisierte signifikante Unterschied zwischen der menschenähnlich- und maschinell-kooperierenden Versuchsbedingung erkennen.

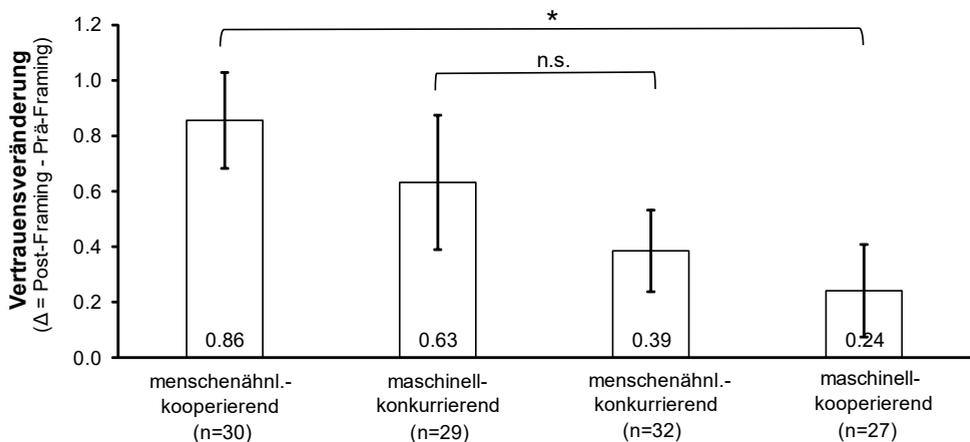


Abbildung 4.18: Grafische Darstellung der Vertrauensveränderung zwischen beiden Messzeitpunkten je Versuchsgruppe.

Der Vertrauenszuwachs in der menschenähnlich-kooperierenden Versuchsbedingung übertraf mit $M = 0.86$ die Steigerung in der maschinell-kooperierenden Versuchsbedingung mit $M = 0.24$ deutlich. Der Unterschied zwischen der Vertrauensentwicklung in der maschinell-konkurrierenden Versuchsbedingung mit $M = 0.63$ und der menschenähnlich-konkurrierenden Versuchsbedingung mit $M = 0.39$ fiel indes deutlich geringer und nicht signifikant aus.

4.5.3.6 Einfluss der UVs auf Misstrauen

Analog zur Analyse des Vertrauens wurde der Einfluss des Framings auf das Misstrauen in den Cobot zunächst mithilfe einer mehrfaktoriellen gemischten $2 \times 2 \times 2$ -ANOVA untersucht. Dabei zeigte sich ein signifikanter Haupteffekt des Messzeitpunktes auf die Misstrauenswerte, $F_{\text{Messzeitpunkt}}(1,114) = 44.20, p < .001, \eta^2 = .279$. Wenngleich die Analyse der Misstrauenswerte keinen signifikanten Interaktionseffekt hervorbrachte, $F_{\text{Interaktion}}(1,114) = 3.53, p = .063, \eta^2 = .030$, zeigt sich auf den Interaktionsdiagrammen in Abbildung 4.19 und Abbildung 4.20 ein Verlaufsmuster, das demjenigen bei der vertrauensbasierten Auswertung ähnelt und eine disordinale Interaktion andeutet.

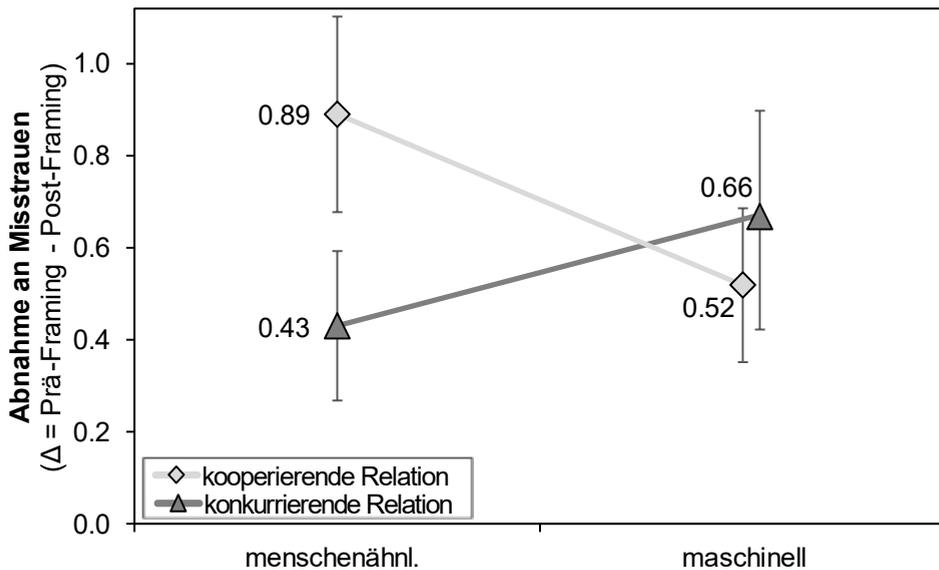


Abbildung 4.19: Interaktionsdiagramm zur Misstrauensveränderung zwischen den beiden Messzeitpunkten (mit wahrgenommener Menschenähnlichkeit abgetragen auf der horizontalen Achse).

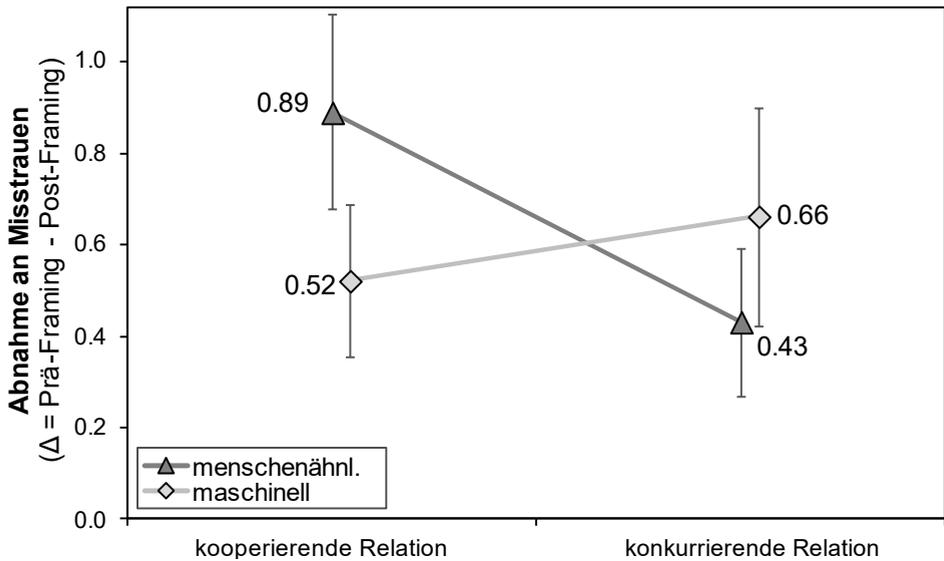


Abbildung 4.20: Interaktionsdiagramm zur Misstrauensveränderung zwischen den beiden Messzeitpunkten (mit wahrgenommener Kooperationsorientierung abgetragen auf der horizontalen Achse).

Abbildung 4.21 stellt ergänzend die Abnahme an Misstrauen pro Versuchsgruppe zwischen den beiden Messzeitpunkten dar.

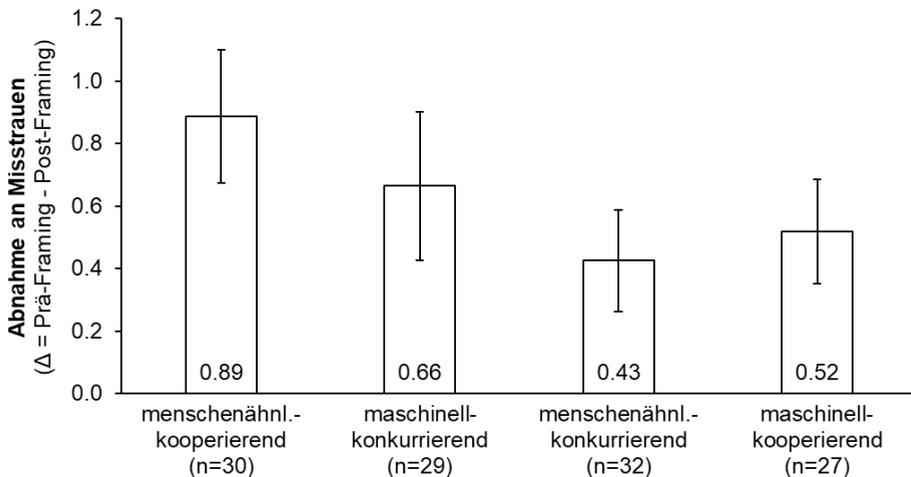


Abbildung 4.21: Grafische Darstellung der Misstrauensabnahme zwischen beiden Messzeitpunkten je Versuchsgruppe. Alle Unterschiede zwischen den Gruppen n. s. ($ps > .093$).

Wie *post hoc-t*-Tests zeigten, liegen keine signifikanten Unterschiede zwischen den Versuchsgruppen vor ($ps > .093$). Konträr zur Untersuchung des Vertrauens bestand die höchste – wenngleich nicht-signifikante – mittlere Differenz zwischen einem menschenähnlich-kooperierenden Cobot mit $M = 0.89$ im Vergleich zu einem menschenähnlich-konkurrierenden Cobot mit $M = 0.43$, $t(114) = 1.69$, $p = .093$, $\eta^2 = .025$.

4.5.3.7 Auswertung der Akzeptanz-Items

Abbildung 4.22 stellt die durchschnittliche Zustimmung zu den abgefragten Akzeptanz-Items dar. Dabei zeigten sich insgesamt nur geringe Abweichungen zwischen den verschiedenen Items, deren Zustimmungswerte sich stark um den Skalenmittelpunkt konzentrierten.

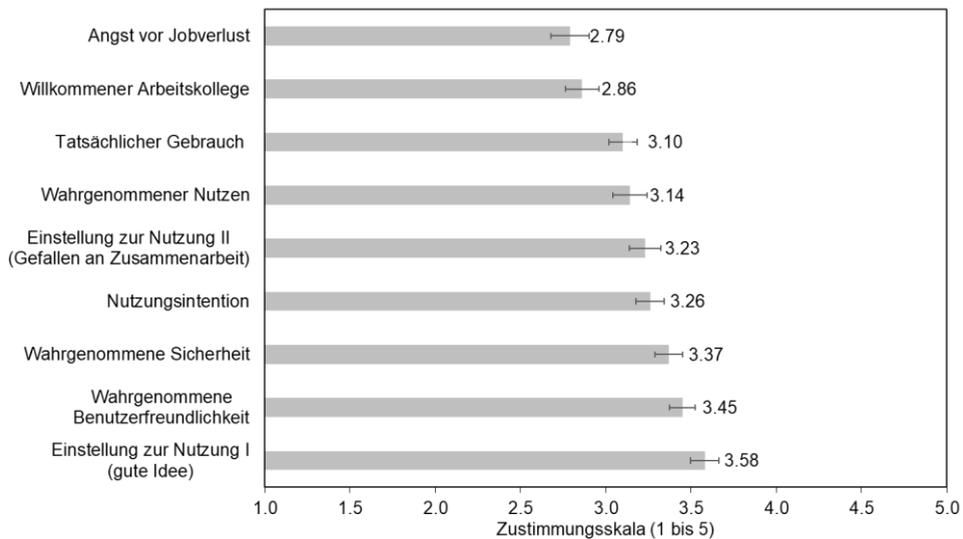


Abbildung 4.22: Deskriptivstatistischer Vergleich der durchschnittlichen Zustimmungswerte zu den abgefragten Akzeptanz-Items.

Die höchste Zustimmung erreichte die Aussage, dass es sich bei der Benutzung des Cobots um eine gute Idee handelt ($M = 3.58$, $SD = 1.02$), zumal dieser als benutzungsfreundlich ($M = 3.45$, $SD = 0.91$) und sicher ($M = 3.37$, $SD = 0.98$) wahrgenommen wurde. Geringere Zustimmungswerte erzielte die Aussage, dass es sich um einen willkommenen Arbeitskollegen handelt ($M = 2.86$, $SD = 1.19$). Außerdem waren die Versuchsteilnehmenden im Mittel *teils-teils* davon überzeugt, dass sie durch den Cobot ihren Job verlieren könnten ($M = 2.79$, $SD = 1.35$). Dieser Wert erscheint vor dem Hintergrund der in den Vorstudien offenkundigen Angst um Arbeitsplatzverlust vergleichsweise niedrig. Allerdings ist bei der Interpretation zu beachten, dass die Versuchsteilnehmenden diese Einschätzung nach dem

Framing vornahmen, welches diese Ängste möglicherweise bereits abgemildert hatte. Ferner deutet die relativ hohe Standardabweichung auf stark ambivalente Haltungen zu dieser Frage hin. Insgesamt sahen sich die Versuchsteilnehmenden eher bedroht, wenn der Cobot menschenähnlich umschrieben wurde. In der Versuchsgruppe mit menschenähnlichem Framing lagen die Mittelwerte im Falle einer kooperierenden Relation bei $M = 2.93$, $SD = 1.46$ anstatt bei $M = 2.33$, $SD = 1.18$ und im Falle einer konkurrierenden Relation bei $M = 3.12$, $SD = 1.26$ anstatt bei $M = 2.86$, $SD = 1.46$ in der Versuchsgruppe mit maschinellem Framing.

4.5.3.8 Explorative Korrelationsanalyse

Eine explorative Korrelationsanalyse verdeutlichte eine signifikante schwach positive Korrelation zwischen der wahrgenommenen Menschenähnlichkeit und der wahrgenommenen Kooperationsorientierung, $r = .207$, $p = .011$, die darauf hinweist, dass die beiden Konstrukte nicht vollständig unabhängig voneinander sind. Während eine menschenähnliche Wahrnehmung des Cobots positiv mit der Intensität der Auseinandersetzung mit dem Thema Robotik korreliert, $r = .189$, $p = .021$, zeigte sich ein statistischer Zusammenhang zwischen einer weniger negativen Einstellung gegenüber Robotern (NARS) und der wahrgenommenen Kooperationsorientierung, $r = -.380$, $p < .000$. In Hinblick auf das Alter der Versuchsteilnehmenden zeigten sich mit Ausnahme der Berufserfahrung keine signifikanten Zusammenhänge mit anderen Variablen, $r = .490$, $p < .001$. Vertrauen und Misstrauen korrelierten sowohl vor dem Framing, $r = -.508$, $p < .001$, als auch nach dem Framing, $r = -.535$, $p = .004$, signifikant stark negativ, was auf einen engen Zusammenhang beider Konstrukte hindeutet.

Das Vertrauen in den Cobot hängt vor dem Framing stark, $r = -.520$, $p < .000$, und nach dem Framing mittelstark, $r = -.342$, $p < .000$, von der grundsätzlichen Einstellung gegenüber Robotern (NARS) ab. Diese Einstellung korreliert wiederum gering mit dem Vorwissen über Roboter, $r = -.264$, $p = .001$, sowie mit der Technikkompetenzüberzeugung, $r = -.281$, $p = .001$. Damit scheint sich zu bestätigen, dass Personen, die sich bereits stärker mit Robotik auseinandergesetzt haben und über eine höhere Technikkompetenzüberzeugung verfügen, auch positiver gegenüber Robotern eingestellt sind und ihnen in höherem Maße vertrauen. Die weniger starke Korrelation nach dem Framing deutet darauf hin, dass die unterschiedliche experimentelle Manipulation als zusätzlicher Einflussfaktor auf das Vertrauen die initialen Zusammenhänge abschwächte.

Ergänzend wurden die Korrelationen zwischen dem Vertrauen nach der Interaktion und den jeweiligen Akzeptanz-Items untersucht, um den Zusammenhang zwischen den beiden oft synonym verwendeten Konstrukten des Vertrauens und der Akzeptanz näher zu explorieren. Tabelle 26 zeigt, dass sämtliche Akzeptanz-Items signifikante mittlere bis starke Korrelationen mit dem Vertrauen und dem Misstrauen nach dem Framing aufweisen. Die Daten legen die Vermutung nahe, dass Vertrauen und Akzeptanz gerade in einem solchen

anwendungsbezogenen Setting sehr ähnliche Konstrukte messen. Aus methodischer Sicht ist zu bedenken, dass die hohe Stichprobenzahl den Ausweis statistischer Signifikanz begünstigt (Lantz, 2013; Royall, 1986). Ferner lässt sich aufgrund der Vielzahl an Interkorrelationen und dem explorativen Charakter der Analyse nicht ausschließen, dass es sich mitunter um sog. Scheinkorrelationen handelt, die durch den Einfluss weiterer Variablen zustande kommen (Döring & Bortz, 2016, S. 684).

Tabelle 26: Ergebnisse der Korrelationsanalyse zwischen Akzeptanz-Items und Vertrauen bzw. Misstrauen, gemessen jeweils nach dem Framing.

	Vertrauen		Misstrauen	
	<i>r</i>	<i>p</i>	<i>r</i>	<i>p</i>
Nutzungsintention	.429	< .001***	-.392	<.001***
Tatsächlicher Gebrauch	.452	<.001***	-.269	.001**
Willkommener Arbeitskollege	.504	<.001***	-.416	<.001***
Wahrgenommener Nutzen	.421	<.001***	-.369	<.001***
Wahrgenommene Benutzerfreundlichkeit	.462	<.001***	-.313	<.001***
Angst vor Jobverlust	-.309	<.001***	.295	<.001***
Einstellung zur Nutzung I	.551	<.001***	-.552	<.001***
Einstellung zur Nutzung II	.477	<.001***	-.467	<.001***
Wahrgenommene Sicherheit	.518	<.001***	-.557	<.001***

4.5.3.9 Qualitative Analyse der Freitext-Aussagen

Insgesamt zeigte sich in den häufig kurzen Freitext-Antworten ein sehr heterogenes Meinungsbild. Wenngleich die Auswertung der Akzeptanz-Items im Mittel keine überwiegend vorhandene Angst vor Arbeitsplatzverlust durch den Cobot-Einsatz indizierte, schienen einige Versuchsteilnehmende fest davon überzeugt zu sein, dass Roboter grundsätzlich Arbeitsplätze vernichten und nahmen deshalb eine sehr ablehnende Haltung ein, wie die folgenden Zitate exemplarisch illustrieren⁹:

- „ein unangenehmer Kollege, eine Art Normbrecher“ (mensenähnlich-konkurierend, 204)
- „Arbeitsplatzvernichter! Der Roboter wird nie in die Sozialsysteme einzahlen und so das ganze Solidarsystem in unserer Gesellschaft bedrohen“ (maschinell-konkurierend, 398)

⁹ In allen wörtlich wiedergegebenen Freitext-Aussagen wurde die Rechtschreibung zum Zweck der besseren Lesbarkeit korrigiert. Die Angaben in Klammern repräsentieren die Teilnehmenden-ID innerhalb der Studie.

- „Eine Maschine, die Arbeitsplätze vernichtet, Menschen noch mehr unter Leistungs- und Erfolgsdruck setzt und krank macht“ (mensenähnlich-konkurrierend, 543)
- „Er wird sympathisch dargestellt, damit der Mensch ihn als etwas Positives sieht. In Wirklichkeit soll er den Menschen ersetzen und überflüssig machen“ (mensenähnlich-kooperierend, 874)
- „mechanisierte Unterstützung bis zur Abschaffung“ (mensenähnlich-konkurrierend, 1222)
- „Ein mechanisches Gerät ohne Seele, dass den Menschen JOBS kostet“ (maschinell-kooperierend, 3797)
- „Roboter nehmen den Menschen die Arbeit weg“ (mensenähnlich-konkurrierend, 6675)

Gleichwohl wurden Cobots und Roboter im Allgemeinen in einigen Kommentaren differenziert und ambivalent betrachtet, nämlich einerseits als willkommene Arbeitserleichterung, die andererseits zum Verlust des Arbeitsplatzes führen könnte:

- „Bei genanntem Paul ist es schwer (...) vorherzusagen, ob er mir den Arbeitsplatz in Zukunft streitig machen würde oder nicht. Er arbeitet sehr gut (...).“ (mensenähnlich-konkurrierend, 1422)
- „Könnte eine Erleichterung in vielen Berufen (...) und eine Bedrohung von Arbeitsplätzen sein.“ (maschinell-kooperierend, 1505)

Wenige Versuchspersonen bewerteten den Cobot-Einsatz grundsätzlich positiv:

- „Es wird in naher Zukunft nötig sein, dass der Mensch mit dem Roboter (...) zusammenarbeitet. Ich betrachte das als Erweiterung und nicht als Beschneidung.“ (maschinell-kooperierend, 1936)

Eine Versuchsperson hob explizit die Wirkung des Framings und seine emotionale Reaktion darauf hervor:

- „Durch die Erzählung hat Paul für mich menschliche Charakterzüge bekommen und bei der Vorstellung, dass er beim Frühstück ausgeschlossen wird, tat er mir richtig leid. Ich hätte lieber 50 Kollegen wie Paul als einen menschlichen Kollegen. Paul ist mir jetzt schon sehr sympathisch.“ (mensenähnlich-konkurrierend, 841)

Wenige Versuchsteilnehmende thematisierten die geringere soziale Präsenz des Cobots im Vergleich zur Zusammenarbeit mit menschlichen Kolleg:innen:

- „Man hat halt dann keinen Arbeitskollegen mehr, mit dem man mal „quatschen“ kann. Ich finde es sehr unpersönlich.“ (maschinell-kooperierend, 1004)

- „Es ist halt eine Maschine. Ich fühle mich mit Menschen wohler.“ (maschinell-konkurierend, 6422)

4.5.4 Diskussion

Aus den statistischen Ergebnissen lassen sich zwei zentrale Erkenntnisse in Hinblick auf die untersuchten Forschungsfragen ableiten. Erstens ergaben sich Hinweise auf die Wirksamkeit von Framing in Bezug auf die wahrgenommene Menschenähnlichkeit eines Cobots und die wahrgenommene Kooperationsorientierung der Mensch-Cobot-Relation (*FF2*). Zweitens zeigte sich auf Basis der varianzanalytischen Auswertungen eine statistische Evidenz für das Vorliegen eines Interaktionseffekts dieser beiden UVs in Hinblick auf das initiale Vertrauen in einen Cobot (*FF3*). Übergreifend zeigt sich damit, dass sich Framing auf das initiale Vertrauen in einen Cobot auswirkt.

4.5.4.1 Auswirkungen des Framings

Angesichts der geringen Anzahl an Studien, die die Effektivität von sprachlichem Framing im MRI-Kontext untersuchten, bietet die im Rahmen der Experimentalstudie durchgeführte Manipulationskontrolle relevante Hinweise darauf, dass sich sprachliches Framing zumindest unter gewissen Rahmenbedingungen signifikant auf die Wahrnehmung eines Cobots auswirkt. Bereits Texte mit einer Länge von ca. 260 bzw. 150 Wörtern, in denen nur vereinzelte Formulierungen variierten und damit unterschiedliche Hervorhebungsframes repräsentierten, reichten offenbar aus, um die initiale mentale Repräsentation des Cobots und der Relation zu demselben zu verändern. Die Effekte auf die wahrgenommene Menschenähnlichkeit konnten neben der Selbstauskunft mittels *Godspeed*-Skala zusätzlich durch eine linguistische Analyse der Freitext-Kommentare bestätigt werden und spiegelten sich überdies in einzelnen Cobot-Beschreibungen explizit wider. Dass sich auch bei der linguistischen Analyse hochsignifikante Gruppenunterschiede zeigten, untermauert die These, dass Sprache einerseits die Mensch-Roboter-Relation konstituiert, andererseits aber auch vorherrschende mentale Modelle repräsentiert (s. a. Coeckelbergh, 2011) und damit entsprechende Äußerungen ein reichhaltiges Repertoire an Erkenntnismöglichkeiten bereitstellen (Fischer, 2011; Fischer et al., 2012). Die übereinstimmenden Ergebnisse auf Basis beider Messmethoden erhöhen die Reliabilität der empirischen Untersuchungen und adressieren damit die Forderung nach einem verstärkten Rückgriff auf indirekte Messmethoden aus früheren Studien (Bethel & Murphy, 2010; Stenzel, Chinellato, del Pobil et al., 2012; Wullenkord, 2017). Diese Forderung lässt sich anhand der vorliegenden Forschungsergebnisse unterstreichen, da sich größere Effekte auf Basis der linguistischen Analyse als auf Basis der Selbsteinschätzung zeigten. In Anlehnung an die Theorie des dualen Anthropomorphismus lassen sich die unterschiedlichen Effektstärken ferner als Ausdruck eines starken impliziten Anthropomorphismus interpretieren. Die vergleichsweise geringen

Effektstärken bei der *Godspeed*-Skala könnten überdies durch die Gestaltung der Skala selbst bedingt sein. Obwohl es sich dabei um ein gängiges Messinstrument handelt, liegt die Vermutung nahe, dass die verwendeten Items bei solch subtilen experimentellen Manipulationen eine zu geringe Sensitivität aufweisen (Onnasch & Roesler, 2021). Die Skalenendpunkte wie z. B. *biologische Bewegungen* oder *sterblich* wirken im industriellen Kontext mutmaßlich bizarr und nötigen die Versuchsteilnehmenden zu „weitreichenden Zuschreibungen“ (*far attributions*) (L. J. Baker et al., 2018, S. 248), wogegen diese eventuell Widerstände hegen. Überdies handelt es sich bei den Begriffspaaren der *Godspeed*-Skala mitunter nicht um exakte Antonyme, sodass diese nicht die idealtypischen Endpunkte eines Kontinuums abbilden (Kaplan, Sanders & Hancock, 2021).

4.5.4.2 Interaktionseffekt

Im weiteren Verlauf der Auswertungen standen die Auswirkungen der unterschiedlichen mentalen Modelle auf das initiale Vertrauen der teilnehmenden Produktionsmitarbeitenden in den Cobot im Fokus (*FF3*). Aufgrund des festgestellten Interaktionseffekts ist die teilweise vertretene Annahme zurückzuweisen, dass Anthropomorphismus grundsätzlich das Technologievertrauen begünstigt (vgl. Kapitel 3.2.4.3). Die experimentellen Befunde legen eine differenzierte Beurteilung der Auswirkungen von Vermenschlichung nahe. Letztere scheinen durch die von den beteiligten Personen konstruierte Mensch-Cobot-Relation moderiert zu werden, die im Spannungsfeld zwischen technologischer Ergänzung und Ersetzung und angesichts damit zusammenhängender ambivalenter Haltungen von besonderer Relevanz ist. Konkreter gefasst bedeutet dies, dass sich die Auswirkungen von Vermenschlichung auf das Vertrauen in einen konkreten Cobot in Abhängigkeit davon unterscheiden, ob dieser Cobot im gegebenen Anwendungskontext als kooperierende Unterstützung oder als bedrohliche Konkurrenz in Hinblick auf den eigenen Arbeitsplatz wahrgenommen wird. Insofern legt die vorliegende Studie als eine der ersten ein umfassenderes und kontextspezifisches Konzept von Vertrauen zugrunde. Im Rahmen dieses Konzepts wird Vertrauen nicht nur handlungsbasiert als die Bereitschaft operationalisiert, sich auf einen Roboter zu verlassen, sondern als Einstellung gegenüber einem Roboter verstanden, die auch weiterführende kontextbezogene Komponenten adressiert. Allen voran ist in diesem Zusammenhang das Vertrauen in die Annahme zu nennen, den eigenen Arbeitsplatz ungeachtet der Einführung des Cobots behalten zu können. An dieser Stelle wird deutlich, dass Vertrauen am Arbeitsplatz verschiedene Bezugsobjekte kennt. Der Begriff des Bezugsobjekts beschreibt dabei, *worauf* konkret vertraut wird, d. h. auf welche Annahme oder Zukunftserwartung sich das Vertrauen bezieht. Der Erhalt des eigenen Arbeitsplatzes als Bezugsobjekt des Vertrauens scheint vor dem Hintergrund der virulenten Ängste vor Arbeitsplatzverlust, die sich in den Vorstudien bereits klar abzeichneten, von besonderer Relevanz zu sein. Auch die qualitative Auswertung der Freitext-Antworten reflektiert erneut die Relevanz und Brisanz der Angst vor dem Arbeitsplatzverlust, die in vielen Kommentaren thematisiert wurde und sich in den statistischen Ergebnissen

abbildete, obwohl die verwendete TAS als Operationalisierung für Vertrauen und Misstrauen keine Items enthält, die sich dezidiert auf den Arbeitsplatzkontext beziehen (vgl. Kapitel 4.5.2.3). Teilweise scheint diese Angst ursächlich für eine kategorische und in harschen Worten vorgetragene Ablehnung des Cobots zu sein. Außerdem zeigte sich mitunter ein starkes Misstrauen gegenüber dem Wahrheitsgehalt positiver Beschreibungen, die als unaufrichtiger Vorwand interpretiert wurden. Der beschriebene Interaktionseffekt ergab sich deskriptiv und weniger stark ausgeprägt auch in Hinblick auf die Entwicklung des Misstrauens, das demnach etwas robuster gegen sprachliche Beeinflussungen zu sein scheint.

Anhand der durchgeführten *post hoc-t*-Tests lässt sich der identifizierte Interaktionseffekt genauer ausbuchstabieren. Ein menschenähnliches wirkt im Vergleich zu einem maschinellen Framing des Cobots positiver auf das Vertrauensniveau, sofern die Mensch-Cobot-Relation als kooperierend angesehen wird. Um diesen Effekt genauer verstehen zu können, ist an dieser Stelle ein Rückgriff auf die zugrundeliegende Vertrauensdefinition sowie eine genauere Betrachtung der vermuteten Einflussfaktoren und Dimensionen des Vertrauens am Arbeitsplatz notwendig. Gemäß der angewandten Vertrauensdefinition nach J. D. Lee und See (2004) lässt sich Vertrauen als die Einstellung verstehen, dass der Cobot bei der Erreichung der Ziele der vertrauenden Person in einer Situation der Unsicherheit und Vulnerabilität hilfreich ist. Entsprechend hängt das Vertrauen in einen Cobot von den konkreten Zielen ab, die eine Person verfolgt und die sich wiederum aus den typischen arbeitsplatzbezogenen Vulnerabilitäten ableiten. Auf Basis der Erkenntnisse aus den Vorstudien und in Anlehnung an Elemente der Akzeptanzmodelle (jeweils in Klammern; vgl. Bröhl et al., 2019) lassen sich die folgenden zentralen Ziele ableiten: körperliche Unversehrtheit (*perceived safety*), Erhalt des Arbeitsplatzes (*ethical implications*), Vermeidung sozialer Isolation (*social implications*) und Freude bei der Arbeit (*perceived enjoyment*), Erzielen guter Arbeitsergebnisse (*output quality*) und Erhalt des Selbstwertes durch die Erfahrung der Wertschätzung eigener Arbeitsleistung (*job relevance* und *self-efficacy*). Tabelle 27 listet diese Ziele sowie die damit verbundenen komparativen Vorteile eines menschenähnlich bzw. maschinell wahrgenommenen Cobots im Falle einer kooperationsorientierten Mensch-Cobot-Relation auf. Einschränkend ist dabei zu erwähnen, dass sich die Einschätzungen zu den jeweiligen komparativen Vorteilen nicht aus den quantitativen experimentellen Daten ableiten lassen, sondern den Status theoriebasierter Hypothesen haben, die der möglichen Erklärung der gefundenen Effekte dienen, aber ihrerseits einer dezidierten empirischen Überprüfung bedürfen. Insbesondere erlauben die qualitativen Einschätzungen keine quantitative Aufsummierung zu einem übergreifenden Vertrauenswert.

Tabelle 27: Ziele am Arbeitsplatz und damit verbundene Auswirkungen in Hinblick auf das Vertrauen in einen menschenähnlich oder maschinell wahrgenommenen Cobot, zu dem eine kooperationsorientierte Relation besteht.

Ziele am Arbeitsplatz	Vorteile im Fall einer kooperationsorientierten Mensch-Cobot-Relation	
	Menschenähnliche Wahrnehmung	Maschinelle Wahrnehmung
Körperliche Unversehrtheit	Vorhandenes Vorwissen über Menschen kann abgerufen und auf den Cobot angewendet werden, was zu einer höheren wahrgenommenen Effektivität und einem geringeren wahrgenommenen Risiko unvorhergesehener, potenziell gefährlicher Aktionen führt.	
Erhalt des Arbeitsplatzes	<i>– irrelevant, da kooperationsorientierte Relation –</i>	
Soziale Interaktion / Spaß bei der Arbeit	Menschenähnliche Beschreibung begründet Annahme eines quasi-sozialen Gegenübers.	
Erzielen guter Arbeitsergebnisse	Höhere wahrgenommene Effektivität im Umgang mit einem vermenschlichten Cobot lässt bessere kooperative Arbeitsergebnisse vermuten, allerdings könnte die Leistungsfähigkeit eines maschinell repräsentierten Cobots aufgrund des PAS (vgl. Kapitel 3.2.4.3) höher eingeschätzt werden.	
Erhalt des Selbstwertes	<i>– irrelevant, da kooperationsorientierte Relation –</i>	

Die Tabelle verdeutlicht, dass in Bezug auf zwei der drei genannten relevanten Ziele spezifische Vorteile eines menschenähnlich wahrgenommenen Cobots bestehen. Der Erhalt des Arbeitsplatzes und des Selbstwertes werden in diesem Szenario als irrelevant eingeschätzt, weil eine kooperationsorientierte Relation keine Wettkampfgedanken und Verlustängste impliziert. Die Argumentation für die aufgezeigten Vorteile ergibt sich aus der Dreifaktoren-Theorie (vgl. Kapitel 3.2.3.3). Demnach führt die Übertragung mentaler Modelle vom Menschen auf den Cobot zu einem Gefühl der Vertrautheit sowie zu einer höheren wahrgenommenen Selbstwirksamkeit im Umgang mit demselben. Eine vermutete bessere Vorhersagbarkeit der Aktionen des Cobots sowie eine höhere wahrgenommene Effektivität mindern Sorgen vor Verletzungen durch unvorhergesehene Bewegungen des Cobots und

steigern die Aussicht auf die Erzielung guter Arbeitsergebnisse im Rahmen einer gelungenen Kooperation mit dem Cobot. Letztere könnte allerdings durch eine höhere wahrgenommene Fehleranfälligkeit von Menschen im Vergleich zu vermeintlich perfekt funktionierenden Maschinen gemindert werden. Ferner begründet eine menschenähnliche Wahrnehmung des Cobots die Aussicht auf die Befriedigung der sozialen Bedürfnisse. Frühere Studien verdeutlichten, dass insbesondere die Sozialitätsmotivation einen einflussreichen Auslöser für Anthropomorphisierung darstellt (Im Shin & Kim, 2018, S. 8). Übereinstimmend mit Untersuchungen von C. Edwards, Edwards, Spence und Westerman (2016) hoben einige Versuchsteilnehmende in den ausgewerteten Freitext-Kommentaren die Wichtigkeit der sozialen Präsenz am Arbeitsplatz hervor, die wiederum mit menschlichen Kolleg:innen anstatt mit Robotern assoziiert wird. Zwar erscheint die Assoziation eines stärker menschenähnlich wahrgenommenen Cobots mit einem sozialen Gegenüber als sehr stereotypisch. Allerdings sind solche vereinfachenden Festlegungen – nicht zuletzt in Anlehnung an die Vertrauensbeschreibung nach Luhmann (1968/2014) (vgl. Kapitel 3.4.1) – geradezu prototypisch für Vertrauensbildungsprozesse. Solche vereinfachenden Assoziationen vermögen es, in Situationen von Unsicherheiten die möglichen Zukunftsentwürfe künstlich zu reduzieren. Davon unbenommen bleibt an dieser Stelle gleichwohl die Frage, inwiefern diese subjektive Erwartungshaltung von der menschenähnlich repräsentierten Maschine erfüllt werden kann. Dieser Aspekt fließt allerdings erst in die dynamische Vertrauenskomponente während einer realen MRI ein, wenn sich die Fähigkeiten des Cobots und dessen Zuverlässigkeit in Form von realen Erfahrungswerten abbilden.

Zusätzlich könnte der statistisch beobachtbare besonders geringe Vertrauenszuwachs in der maschinell-kooperierenden im Vergleich zur menschenähnlich-kooperierenden Versuchsbedingung durch die kognitive Dissonanz (Festinger, 1957) bedingt sein, die entsteht, wenn eine Maschine als unterstützende:r Kooperationspartner:in umschrieben wird, aber mentale Modelle vorherrschen, die Maschinen als eigenständig arbeitende, nicht-interaktive Automatisierungsinstrumente repräsentieren (vgl. Kapitel 4.3.3). Während es Mitarbeitende gewohnt sind, mit unterschiedlichen Kolleg:innen jeweils eher in kooperations- oder in konkurrenzorientierten Beziehungen zu stehen, liegt es nahe, dass bekannte Maschinen traditionell im Kontext mentaler Modelle von Konkurrenz repräsentiert werden bzw. kein mentales Modell für eine kooperationsorientierte Relation besteht. Eine Beschreibung einer Maschine als Kolleg:in i. e. S. würde somit zu konfligierenden Kognitionen führen. Hierbei handelt es sich allerdings um hypothetische Begründungen, zumal der Einfluss von kognitiver Dissonanz in Mensch-Roboter-Relationen bisher wenig untersucht wurde (Levin, Harriott, Paul, Zheng & Adams, 2013).

Auf Basis des signifikanten Interaktionseffektes liegt die Annahme nahe, dass umgekehrt bei Vorliegen einer konkurrenzorientierten Relation ein maschinelles Framing eine vergleichsweise positivere Vertrauensveränderung als ein menschenähnliches Framing dieses

Cobots zur Folge hat. Diese Hypothese ließe sich inhaltlich dadurch begründen, dass der eigene Selbstwert bei einer antizipierten Ersetzung durch einen Menschen stärker bedroht wird als bei der Ersetzung durch eine Maschine (Granulo et al., 2019). Inferenzstatistisch konnte dieser Zusammenhang in den *post hoc t*-Tests allerdings nicht nachgewiesen werden. Wie Tabelle 28 analog zu Tabelle 27 darstellt, stehen dem beschriebenen Einfluss auf den Selbstwert gegenläufige Tendenzen gegenüber, die in Summe den postulierten Effekt möglicherweise unterminieren. Im Vergleich zu Tabelle 27 wird deutlich, dass einige Vorteile eines menschenähnlichen Cobots auch im Falle einer konkurrenzorientierten Relation bestehen bleiben.

Tabelle 28: Ziele am Arbeitsplatz und damit verbundene Auswirkungen in Hinblick auf das Vertrauen in einen menschenähnlich oder maschinell wahrgenommenen Cobot, zu dem eine konkurrenzorientierte Relation besteht.

Ziele am Arbeitsplatz	Vorteile im Fall einer konkurrenzorientierten Mensch-Cobot-Relation	
	Menschenähnl. Wahrnehmung	Maschinelle Wahrnehmung
Körperliche Unversehrtheit	Vorhandenes Vorwissen über Menschen kann abgerufen und auf den Cobot angewendet werden, was zu einer höheren wahrgenommenen Effektivität und einem geringeren wahrgenommenen Risiko unvorhergesehener, potenziell gefährlicher Aktionen führt.	
Erhalt des Arbeitsplatzes	Grad der wahrgenommenen Bedrohung des Arbeitsplatzes hängt davon ab, ob Menschen oder Maschinen als kompetenter für die jeweilige Aufgabe eingeschätzt werden; hierzu existieren allerdings konfligierende experimentelle Ergebnisse, vgl. Waytz et al. (2014) im Gegensatz zu J. J. Choi & Kwak (2015) und Goetz et al. (2012).	
Soziale Interaktion / Spaß bei der Arbeit	– <i>irrelevant, da kein Interesse an sozialer Interaktion mit Konkurrent:innen</i> –	
Erzielen guter Arbeitsergebnisse	Höhere wahrgenommene Effektivität im Umgang mit einem vermenschlichten Cobot lässt bessere Arbeitsergebnisse vermuten, sofern in der konkurrenzorientierten Relation überhaupt eine Kooperationsbereitschaft gegeben ist. Wenn dies nicht der Fall ist, wäre davon auszugehen, dass die Einschätzung der Qualität der eigenen Arbeitsergebnisse	

weitestgehend unabhängig von der Wahrnehmung des Cobots ausfällt und das Kriterium damit irrelevant ist.

Erhalt des Selbstwertes

Selbstwert wird in geringem Maße beschädigt, da Maschinen als prinzipiell inkommensurabel zu Menschen eingeschätzt werden (Granulo et al., 2019).

Die Sozialitätsmotivation spielt im Szenario einer konkurrenzorientierten Relation eine untergeordnete Rolle, da von einem geringen Interesse an sozialen Interaktionen mit Konkurrent:innen auszugehen ist. Gleichzeitig gewinnt die vermeintliche Vertrautheit mit dem Gegenüber an Bedeutung, die sich laut Drei-Faktoren-Theorie aus der Anwendung mentaler Modelle vom Menschen ergibt. Diese ist innerhalb einer kompetitiven Relation insofern besonders relevant, als sich aus präzisen Einschätzungen der komparativen Leistungsfähigkeit einer konkurrierenden Technologie die Wahrscheinlichkeit einer Ersetzung durch dieselbe ableiten lässt. Dieser Faktor allein spräche für ein höheres Vertrauen in einen menschenähnlich wahrgenommenen Cobot im Konkurrenzszenario. Dass sich in den deskriptiven Daten allerdings eine höhere – wenngleich nicht-signifikante – Vertrauenssteigerung beim maschinellen Cobot abzeichnet, spricht dafür, dass die Vorstellung einer möglichen Ersetzung durch einen maschinell wahrgenommenen im Vergleich zu einem menschenähnlich wahrgenommenen Cobot den Selbstwert weniger stark beschädigt. Dabei ist zu berücksichtigen, dass diese Argumentation auf einer Studie von Granulo et al. (2019) basiert, die die antizipierte Ersetzung durch eine Maschine der Ersetzung durch einen Menschen gegenüberstellte. In der durchgeführten Experimentalstudie stand in allen Versuchsbedingungen die Ersetzung durch einen Cobot im Raum. Lediglich das Framing unterschied sich und beeinflusste damit die mentale Repräsentation in einem gewissen Umfang. Die deskriptiven Werte bestätigen allerdings, dass auch der menschenähnlich geframte Cobot weiterhin eher als Maschine denn als Mensch wahrgenommen wurde. Das Auftreten ähnlicher starker Effekte wie in der Studie von Granulo et al. (2019) wäre überraschend gewesen angesichts der Tatsache, dass in der vorliegenden Studie lediglich die Wahrnehmung eines Gegenübers in geringem Umfang variiert wurde. Angesichts dieser deutlichen Diskrepanz im experimentellen Design lassen sich selbst vergleichsweise kleine Effekte als Hinweis darauf werten, dass die Psychologie der technischen Ersetzung einen relevanten Einflussfaktor auf das Vertrauen darstellt. Diese im Vergleich zum kooperierenden Szenario geringere reale Effektstärke könnte angesichts einer gleichbleibenden Stichprobengröße zu insignifikanten statistischen Ergebnissen bei den *post hoc-t*-Tests im konkurrierenden Szenario geführt haben. Dies scheint insofern plausibel, als eine partielle *a posteriori*-Teststärkenanalyse (Dziak, Dierker & Abar, 2020) mit *G*Power* (Version 3.1.9.4) ergab, dass selbst für den theoretisch vermeintlich größeren Effekt von $\eta^2 = .045$,

der sich beim Vergleich der menschenähnlich- und maschinell-kooperierenden Versuchsbedingung ergab, eine Stichprobe von $n = 174$ nötig gewesen wäre, um eine akzeptable Teststärke von $(1-\beta) = 80\%$ zu erreichen.

4.5.4.3 Versuchsgruppenunabhängiger Vertrauensanstieg

Ferner zeigte sich in der Auswertung ein Vertrauensanstieg zwischen beiden Messzeitpunkten in allen Experimentalbedingungen, wobei dieser Anstieg in drei Bedingungen signifikant ausfiel. Auf Basis des wesentlich schwächeren und nicht-signifikanten Vertrauenszuwachses in der Kontrollgruppe lässt sich eine rein methodische Erklärung aufgrund eines *test-retest bias* ausschließen. Offenbar scheint die bloße Konfrontation mit Informationen über einen konkreten Cobot einen Effekt auf das Vertrauensniveau zu erwirken, unabhängig davon, wie genau die Informationen sprachlich gerahmt sind. Dieser Umstand lässt sich durch den sog. *Effekt des bloßen vorgestellten Kontakts* (*mere imagined exposure effect*) nach Miles und Crisp (2014, S. 20) erklären. Frühere Studien haben bereits den positiven Effekt von Vorerfahrungen in der Interaktion mit Robotern auf das Vertrauensniveau nachgewiesen (Sanders et al., 2017) und den aus der Sozialpsychologie bekannten sog. *Effekt der Darbietungshäufigkeit* (*mere exposure effect*; Bornstein, 1989) damit in Verbindung gebracht (Wullenkord, Fraune, Eyssel & Šabanović, 2016). Bemerkenswerterweise offenbarten weitere Studien ebenfalls einen Rückgang negativer Empfindungen gegenüber einem Roboter, wenn sich Personen den Kontakt mit diesem lediglich vorstellten (Wullenkord et al., 2016, S. 984). So führte ein für wenige Minuten andauernder imaginiertes Kontakt mit einem unbekanntem Roboter im Sinne der mentalen Simulation einer Interaktion zu einer höheren Akzeptanz und zu einer weniger negativen Einstellung diesem Roboter gegenüber, insbesondere wenn ein kooperatives Szenario imaginiert wurde und wenn sich das imaginierte und reale Interaktionsszenario möglichst glichen (Kuchenbrandt & Eyssel, 2012, S. 466f.; Wullenkord, 2017, S. 156; Wullenkord & Eyssel, 2019). Das lässt sich unter Rückgriff auf die Skripttheorie (Abelson, 1981) damit erklären, dass Situationen weniger Unsicherheit und mehr Wohlgefühl auslösen, wenn sie gemäß vorher aktivierter stereotyper Verhaltensskripte ablaufen. Möglicherweise spielt überdies eine Rolle, dass mentale Simulationen neuronale Verknüpfungen aktivieren, die mit dem realen Erleben der Handlung assoziiert sind und entsprechende Bewegungsabläufe bereits vorplanen, sodass hinsichtlich der neuronalen Verarbeitung deutliche Ähnlichkeiten zwischen der bloßen Vorstellung und dem realen Erleben einer Situation bestehen (Wehling, 2016, S. 21f.). Die handlungsorientierten Auswirkungen des Effekts der imaginierten Interaktion sind allerdings noch unklar (Wullenkord & Eyssel, 2019).

Wenn Personen Texte über Roboter lesen, die sprachlich in gewisse Deutungsrahmen eingebettet sind, könnte das dazu führen, dass sich die Rezipienten eine Interaktion mit einem Roboter vorstellen, insbesondere wenn sie noch über wenige Verhaltensskripte für solche Interaktionen verfügen. Zwar wurden die Versuchsteilnehmenden nicht explizit dazu

aufgefordert, eine Interaktion mit dem Cobot zu imaginieren, allerdings liegt die Vermutung nahe, dass dies bei der Lektüre der Framing-Texte über alle Experimentalbedingungen hinweg geschah. In der Kontrollgruppe, die keine Informationen über Roboter erhielt, ließ sich kein vergleichbarer vertrauenssteigernder Effekt feststellen. Insofern beinhaltet jedwede Intervention des sprachlichen Framings möglicherweise einen nicht-intendierten Nebeneffekt. Die Informationsbereitstellung führt zu einem Gefühl der Informiertheit, das das Zutrauen in die eigene Urteilsfähigkeit steigert und zu einer mentalen Simulation einer künftigen Interaktion führt. Diese stellt bereits die Festlegung auf ein zukünftiges Verhaltensskript dar und impliziert damit eine vertrauenssteigernde Disambiguierung der bevorstehenden Interaktionssituation. Ferner thematisiert die aktuelle Forschung inzwischen verstärkt die Bereitstellung sog. Transparenzinformationen (*transparency information*) zur Steigerung des Zutrauens in die eigene Urteilsfähigkeit. Die Grundannahme lautet dabei, dass mehr verfügbare Informationen über die Funktionsweise eines Roboters das Vertrauen in denselben begünstigen, da sie dessen Innenleben für externe Beobachtende einsehbar, also transparent machen (Malle, Fischer, Young, Moon & Collins, 2020, S. 3; Miller, 2021, S. 240). Dabei können die Transparenzinformationen vor oder während der Interaktion entweder von Dritten wie z. B. der Unternehmensleitung bei einer betrieblichen Robotereinführung (Wurhofer et al., 2015) oder durch den Roboter selbst während der Interaktion (Sanders, Wixon, Schaefer, Chen & Hancock, 2014) bereitgestellt werden.

Imaginierter Kontakt könnte eine Interventionsmöglichkeit darstellen, um Menschen „psychologisch auf eine bevorstehende MRI vorzubereiten“ und die Akzeptanz bei neuen Roboternutzenden zu erhöhen bzw. die Angst vor Robotern zu mindern (Wullenkord & Eyszel, 2014, S. 493; Wullenkord, 2017). Dieser Effekt wird gemeinhin als Alternativlösung betrachtet, wenn realer Kontakt unmöglich ist. Aktuelle qualitative Studien mit Produktionsmitarbeitenden untermauern den Eindruck, dass die Art und Weise, wie Mitarbeitende über eine Cobot-Einführung informiert werden, zu einem erheblichen Maße auf Vertrauen und Akzeptanz wirkt (Meißner et al., 2020). Die experimentellen Ergebnisse bekräftigen insofern die These, dass bereits die Imagination eines zukünftigen Kontakts auf Basis vermittelter Informationen über einen Cobot einen vertrauensfördernden Effekt auszulösen vermag (Kuchenbrandt & Eyssel, 2012; Wullenkord & Eyssel, 2014; Wullenkord et al., 2016; Wullenkord, 2017; Wullenkord & Eyssel, 2019). Allerdings ist davon auszugehen, dass dieser Effekt nur dann auftritt, wenn nicht ohnehin durch relevante Vorkenntnisse oder reale Erfahrungswerte mit MRI-Situationen bereits vorgefertigte mentale Modelle und Verhaltensskripte für die entsprechende Situation vorliegen. Weitere Studien sind erforderlich, um den Effekt genauer zu explorieren und empirisch abzusichern.

4.5.4.4 Zusammenhänge zwischen Konstrukten

In der explorativen Korrelationsanalyse verdeutlichten signifikante Zusammenhänge zwischen Vertrauen, Misstrauen und gemessenen Akzeptanz-Items, dass diese Konstrukte in

der praktischen Verwendung nur schwer zu differenzieren sind und in der Alltagssprache häufig synonym verwendet werden. Ferner zeigt sich in Einklang mit der Literatur (vgl. Müller-Abdelrazeq, 2020, S. 38f.), dass Faktoren wie die Grundeinstellung gegenüber Robotern, die Technikkompetenzüberzeugung sowie das (theoretische) Vorwissen über Roboter das initiale Vertrauen stark beeinflussen. Diese Faktoren hängen wiederum miteinander zusammen. So geht eine intensivere Auseinandersetzung mit Technik und eine stärkere Kontrollüberzeugung in Bezug auf Roboter mit einer positiveren Grundeinstellung einher. Das Alter der Versuchsteilnehmenden spielte in der Korrelationsanalyse hingegen keine direkte Rolle, wenngleich mithin kolportiert wird, dass ein höheres Lebensalter eine geringere Technologieakzeptanz zur Folge hat. Diese Behauptung unterschlägt allerdings die Komplexität des Zusammenspiels der beteiligten Persönlichkeitsmerkmale, bei denen konfundierende Variablen zu berücksichtigen sind (Claßen, 2013, S. 269f.). So zeigen Studien, dass sich ältere Versuchspersonen im Durchschnitt weniger über Roboter informierten als jüngere Teilnehmende (MacDorman, Vasudevan & Ho, 2009) und sich weniger kompetent im Umgang mit modernen Computern fühlten (Marquié, Jourdan-Boddaert & Huet, 2002). Möglicherweise setzen sich ältere Menschen weniger stark mit Robotik auseinander und verfügen über eine geringere Kompetenzüberzeugung, wenngleich sich mit zunehmendem Alter und steigender Berufserfahrung die Wahrscheinlichkeit erhöht, praktische Erfahrungen mit Robotern zu erwerben. Somit könnten die Intensität der Beschäftigung mit Robotern sowie die Technikkompetenzüberzeugung Mediatorvariablen darstellen, die einen indirekten Zusammenhang zwischen einem hohen Lebensalter und einem geringeren Vertrauen in eine Technologie bedingen. Einschränkend ist zu erwähnen, dass die Studienlage bisher nicht eindeutig und die Übertragbarkeit auf den industriellen Sektor unklar ist (Henderson, 2015; Müller-Abdelrazeq, 2020, S. 39).

4.5.5 Zwischenfazit

Zusammenfassend lässt sich in Rückbezug zu den Forschungsfragen konstatieren, dass sich mittels sprachlichen Framings die Wahrnehmung der Menschenähnlichkeit eines Cobots und der Relation zu demselben beeinflussen lässt (*FF2*). Die Effektstärken fallen zwar eher gering aus, sind allerdings angesichts der wenig invasiven Intervention anhand zweier kurzer Texte dennoch bemerkenswert. Ferner wirken sich diese beiden Faktoren auf das initiale vermittelte Vertrauen und Misstrauen in einen Cobot aus, wobei die wahrgenommene Kooperationsorientierung der Mensch-Cobot-Relation die Auswirkungen der Vermenschlichung auf das empfundene Vertrauen moderiert (*FF3*). Die in den Vorstudien bereits evidente Angst vor Arbeitsplatzverlust und die sensible Reaktion auf die konstruierte Mensch-Roboter-Relation schlägt sich insofern auch quantitativ messbar nieder und bekräftigt die Notwendigkeit, diesen Aspekt bei Studien im Arbeitsplatzkontext dezidiert einzubeziehen.

Die Versuchsteilnehmenden vertrauten einem menschenähnlich beschriebenen Cobot mehr als einem maschinell beschriebenen, sofern sie sich in einer kooperierenden Relation zu diesem verorteten. Die höheren Vertrauenszuwächse durch das menschenähnliche Framing lassen sich unter Berücksichtigung der vielfältigen Ziele von Mitarbeitenden am Arbeitsplatz auf Basis der Drei-Faktoren-Theorie erklären. Wurde die Mensch-Cobot-Relation als konkurrierend aufgefasst, drehte sich dieser Zusammenhang zwar um, allerdings ohne signifikante Unterschiede aufzuweisen. Der Interaktionseffekt erklärt sich dadurch, dass Menschen ihren Selbstwert eher bedroht sehen, wenn sie eine Ersetzung durch einen Menschen anstatt durch eine Maschine vermuten. Dieses Phänomen ist offenbar auf unterschiedlich wahrgenommene Cobots übertragbar, sodass sich ein höherer Vertrauenszuwachs in der maschinell- im Vergleich zur menschenähnlich-konkurrierenden Versuchsbedingung ergab. Insofern lassen sich kontextspezifische Einflüsse auf das Vertrauen am Arbeitsplatz nur verstehen, wenn dessen verschiedenartige Dimensionen und Bezugsobjekte differenziert betrachtet werden. Die im Rahmen der Diskussion vorgenommene Ausdifferenzierung und Zurechnung einzelner Faktoren auf die Dimensionen ist als rein theoriegeleitet und hypothetisch zu betrachten und bedarf daher weiterer Untersuchungen.

Der beschriebene Interaktionseffekt deutete sich antiproportional auch in Bezug auf Misstrauen an, wobei er hierbei nicht signifikant ausfiel, was auf eine geringere Beeinflussbarkeit von Misstrauen durch Framing hindeutet. Ein weiterer interessanter Befund bestand in einem steigenden Vertrauens- und einem sinkenden Misstrauensniveau zwischen den beiden Messzeitpunkten vor und nach dem Framing. Dass sich signifikante Vertrauenssteigerungen nur in den Experimentalbedingungen, nicht aber in der Kontrollbedingung einstellen, suggeriert einen vertrauenssteigernden Effekt durch die Bereitstellung von Informationen über einen Cobot, unabhängig vom Framing der Informationen. Das lässt sich durch einen Einfluss des imaginierten Kontakts mit einem unbekanntem Cobot erklären, der offenbar einen Nebeneffekt von Framing-Interventionen darstellt.

Da es sich bei den Versuchsteilnehmenden um Produktionsmitarbeitende handelte, ist von einer hohen externen Validität der Studienergebnisse auszugehen. Einschränkend ist zu erwähnen, dass bedingt durch das Online-Setting weder Auswirkungen des Framings auf eine reale Interaktion gemessen noch die Wirkdauer der Framing-Effekte beurteilt werden konnten. Außerdem bleiben affektive Vertrauenskomponenten in Online-Settings weitgehend unberücksichtigt. Ferner wurden gezielt Mitarbeitende ohne praktische Vorerfahrungen mit Robotern ausgewählt, weil sich deren weniger gefestigte mentale Modelle hypothetisch leichter durch sprachliches Framing beeinflussen lassen. Die Übertragbarkeit der Erkenntnisse auf Mitarbeitende mit praktischen Vorerfahrungen und andere Personengruppen sowie die Auswirkungen auf eine reale MRI blieben damit zunächst offene Forschungsfragen, die in der nachfolgenden Experimentalstudie adressiert wurden.

4.6 Experimentalstudie 2: Präsenzstudie mit Studierenden

4.6.1 Fragestellungen und Studiendesign

Angesichts der in Experimentalstudie 1 gewonnenen Indizien für die Auswirkungen von sprachlichem Framing auf das initiale Vertrauen in einen Cobot bestand das Ziel der zweiten Experimentalstudie darin, weiterführende praxisrelevante Fragen zu untersuchen. Diese bezogen sich einerseits auf die Auswirkungen des initialen Vertrauens auf die Wahrnehmung einer realen MRI und andererseits auf die Langfristigkeit von Framing-Effekten:

- *FF4*: Inwiefern wirken sich durch das Framing induzierte Veränderungen am initialen Vertrauen auf die reale Interaktion aus?
- *FF5*: Inwiefern überdauern Framing-Effekte eine reale MRI?

Abbildung 4.23 zeigt schematisch das Design von Experimentalstudie 2, das zunächst demjenigen von Experimentalstudie 1 ähnelt, allerdings die Durchführung einer realen MRI beinhaltet und daher als Präsenzexperiment durchgeführt wurde. Das experimentelle Design wurde um einen dritten Messzeitpunkt des Vertrauens und Misstrauens nach der Durchführung der MRI ergänzt. Die Wahrnehmung der MRI durch die Versuchsteilnehmenden wurde als weitere Variable aufgenommen, die hypothetisch vom Vertrauensniveau vor der Interaktion abhängt. Im Vergleich zu Experimentalstudie 1 wurde auf die Erhebung der Mensch-Cobot-Relation als zweite UV verzichtet. Das liegt darin begründet, dass aufgrund des zeitintensiveren Präsenzversuchs auf Studierende als Stichprobe zurückgegriffen wurde. Da diese i. d. R. noch keinem Beruf nachgehen, erschien die Abfrage von Ängsten vor Jobverlust zu hypothetisch und artifiziell. Überdies war mit einer im Vergleich zu Experimentalstudie 1 geringeren Stichprobenanzahl und dadurch verringerten Teststärke zu rechnen, die durch eine geringe Anzahl an Experimentalgruppen zumindest partiell kompensiert werden sollte.

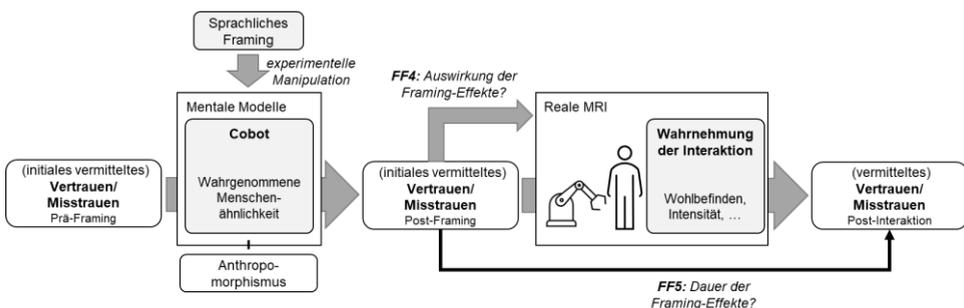


Abbildung 4.23: Schematische Darstellung des Designs von Experimentalstudie 2.

Demnach liegt ein vollfaktorielles gemischtes 2x3-Studiendesign mit dem zweistufigen Zwischensubjektfaktor *wahrgenommene Menschenähnlichkeit* sowie dem dreistufigen Innersubjektfaktor *Messzeitpunkt* mit den Ausprägungen *vor dem Framing*, *nach dem Framing* und *nach der Interaktion* vor. Als quasi-metrische AV wird jeweils die Vertrauensveränderung bzw. die Misstrauensveränderung herangezogen.

4.6.2 Methoden und Material

4.6.2.1 Stichprobenbeschreibung

Insgesamt nahmen 80 Versuchspersonen an der Studie teil. Bei zwölf Versuchsdurchführungen kam es zu Fehlfunktionen des Cobots, die einen Eingriff der Versuchsleitung erforderten. Die entsprechenden Datensätze wurden von der Auswertung ausgeschlossen. Entsprechend basieren die folgenden Auswertungen auf einer Stichprobe von $n = 68$ Teilnehmenden.

Bei allen Studienteilnehmenden handelte es sich um Studierende oder (kürzliche) Absolvent:innen der Hochschule Karlsruhe. Das Alter der Versuchsteilnehmenden lag bei $M = 24.07$ Jahren ($Min = 19$, $Max = 33$, $SD = 2.93$). 24 Teilnehmende waren weiblich (35.3 %) und 44 männlich (64.7 %). Der bisherige höchste Bildungsabschluss der Versuchspersonen war zu 50.0 % das Abitur/die Fachhochschulreife und zu 5.9 % eine abgeschlossene Berufsausbildung. Weitere 44.1% gaben an, bereits einen Hochschulabschluss zu besitzen; dabei handelte es sich in 38.2 % der Fälle um einen Bachelor- und bei 5.9 % um einen Master-Abschluss. Die meisten Umfrageteilnehmenden studierten International Management (38.2 %), Wirtschaftsingenieurwesen (32.4 %) und Wirtschaftsinformatik (16.2 %). Die restlichen Befragten (13.2 %) gaben an, in einem anderen Studiengang zu studieren.

Der Wert für die eigene Technikkompetenzüberzeugung lag durchschnittlich bei $M = 4.04$, $SD = 0.76$ und zeigt, dass die Teilnehmenden *überwiegend* zustimmten, sich im Umgang mit Technik kompetent zu fühlen. Im Durchschnitt hatten sich die Teilnehmenden *mittelmäßig* intensiv mit Robotern auseinandergesetzt, bspw. durch Filme, Zeitungsartikel oder im Internet, $M = 2.75$, $SD = 1.06$. Außerdem zeigte sich eine *mittelmäßig* negative Grundeinstellung der Studierenden gegenüber Robotern, $M = 2.58$, $SD = 0.40$.

4.6.2.2 Material

Als Stimuli für die experimentelle Manipulation wurden die Texte aus Experimentalstudie 1 in geringfügig angepasster Form verwendet. Die Anpassungen waren erforderlich, damit die Beschreibungen aus dem Text nicht im Widerspruch zu den tatsächlichen Fähigkeiten des Cobots standen, der in der realen MRI verwendet wurde. Die angepassten Texte hatten jeweils eine vergleichbare Länge und differierten nur leicht in der Lesbarkeit, die

mittels des Flesch-Lesbarkeitsindex ermittelt wurde. Diese Unterschiede waren erneut durch das Framing bedingt. Tabelle 29 zeigt die Daten zur Länge und Lesbarkeit der jeweiligen Texte. Die vollständigen Texte im Wortlaut finden sich in Anhang C.I und C.II.

Tabelle 29: Länge und Lesbarkeit der verwendeten Framing-Texte in den jeweiligen Versuchsbedingungen (Experimentalstudie 2).

	Wahrgenommene Menschenähnlichkeit	
	Menschenähnliche Beschreibung	Maschinelle Beschreibung
Flesch-Wert	47	56
Lesbarkeit	schwierig	anspruchsvoll
Sätze	19	19
Wörter	271	256
Silben	549	483
Zeichen	1911	1723

4.6.2.3 Messmethodik

Im Wesentlichen wurde die identische Messmethodik wie in Experimentalstudie 1 eingesetzt. Lediglich das in Experimentalstudie 1 über ein einzelnes Item erhobene Konstrukt *Technikkompetenzüberzeugung* wurde in Experimentalstudie 2 zusätzlich mittels einer anerkannten fünfstufigen Multi-Item-Skala bestehend aus vier Items zur Technikkompetenzüberzeugung von Neyer et al. (2012, S. 88) gemessen, um eine bessere Reliabilität zu erzielen. Aus Gründen der Vergleichbarkeit wurde das Single-Item ebenfalls beibehalten. Ferner wurde die Auswahl der Akzeptanz-Items im Vergleich zu Experimentalstudie 1 angepasst. Das resultierende Set an Items ist in Tabelle 30 dargestellt. Ein kompletter Überblick über alle Akzeptanz-Items, die in den beiden Experimentalstudien verwendet wurden, findet sich außerdem in Anhang C.VII.

Tabelle 30: Verwendete Items zur Messung verschiedener Facetten der Akzeptanz (Experimentalstudie 2). Die Items 8 bis 13 wurden auch in Experimentalstudie 1 verwendet.

	Konstrukt	Item	Quelle
1	Selbstwirksamkeit	Nach einer entsprechenden Einweisung wäre ich in der Lage, einen Roboter wie [Robotername] zu benutzen.	TAM 3

	Konstrukt	Item	Quelle
2	Bereitschaft zur Arbeit am gemeinsamen Arbeitsplatz	Es würde mir nichts ausmachen, wenn [Robotername] mit mir an einem gemeinsamen Arbeitsplatz arbeitet.	TAM-HRC als Konstrukt <i>rechtliche Implikation</i>
3	Bereitschaft zum Berichten von Ergebnissen	Ich hätte keine Schwierigkeiten damit, anderen von den Ergebnissen des [Robotername] zu berichten.	TAM 2
4	Ergebnisqualität	Die Ergebnisqualität des [Robotername] wäre hoch.	TAM 2
5	Wahrgenommenes Vergnügen	Ich hätte bei der Nutzung des [Robotername] Spaß.	TAM 3
6	Angst vor sozialer Isolation	Ich würde befürchten, dass ich durch [Robotername] den Kontakt zu meinen Kollegen verliere.	TAM-HRC als Konstrukt <i>soziale Implikation</i>
7	(Generelle) Roboterangst	In der Gegenwart von [Robotername] würde ich mich unwohl fühlen.	TAM 3
8	Wahrgenommene Benutzerfreundlichkeit	Die Interaktion mit [Robotername] wäre einfach.	TAM
9	Wahrgenommener Nutzen	[Robotername] wäre für meine Arbeit nützlich.	TAM
10	Nutzungsintention	Wenn ich wählen könnte, ob mich [Robotername] bei meiner Arbeit unterstützen soll, würde ich mich für die Unterstützung durch [Robotername] entscheiden.	TAM
11	Wahrgenommene Sicherheit	Bei der Nutzung von [Robotername] würde ich mich sicher fühlen.	TAM-HRC
12	Tatsächlicher Gebrauch	Ich würde [Robotername] gegenüber anderen Maschinen im industriellen Umfeld bevorzugen.	TAM
13	Angst vor Jobverlust	Ich würde befürchten, dass ich durch [Robotername] langfristig meinen Job verlieren werde.	TAM-HRC als Konstrukt <i>ethische Implikation</i>

Ergänzend wurden zur Bewertung der realen Interaktion mit dem Cobot die Zustimmungswerte auf einer fünfstufigen Skala zu den folgenden eigens formulierten Items abgefragt¹⁰:

- Während der Interaktion habe ich mich wohl gefühlt.
- Die direkte Zusammenarbeit mit [Paul / dem UR-5] fordert viel Aufmerksamkeit von mir.
- [Paul / Der UR-5] ist mein Kollege.
- [Paul / Der UR-5] passt sich an meine Arbeitsgeschwindigkeit an.
- [Paul / Der UR-5] wartet auf mich, wenn ich mit einem Schritt noch nicht fertig bin.

Außerdem wurden die Versuchsteilnehmenden gebeten, auf einer fünfstufigen Skala von 1 = *gar nicht* bis 5 = *sehr* die Intensität der Interaktion zu bewerten.

Zur impliziten Messung der wahrgenommenen Menschenähnlichkeit wurde wie in Experimentalstudie 1 eine Inhaltsanalyse anhand derselben Kodierleitfäden (vgl. Anhang C.III und C.IV) durchgeführt. Die *inter-rater*-Übereinstimmung lag in Experimentalstudie 2 bei Krippendorffs $\alpha = .949$. Im Vergleich zu Experimentalstudie 1 ergaben sich allerdings geringere Reliabilitätswerte für die verwendeten Skalen anhand von Cronbachs α , die durch die geringere Stichprobenanzahl bedingt sein könnten. Dennoch lagen die Reliabilitätswerte aller verwendeten Skalen auf einem Niveau von $\alpha > .60$. Die genauen Werte sind im Anhang C.VIII.b aufgeführt.

4.6.2.4 Studienablauf

Hinweis: Einige Abbildungen in diesem Kapitel sowie die Beschreibung des Interaktionsszenarios sind stark angelehnt an Baumgartner (2020) (betreute Abschlussarbeit).

Der Studienablauf gliedert sich in Einweisungen und Vorabbefragungen, eine tatsächlich stattfindende MRI und Nachbefragungen sowie ein *debriefing*. Im Vorfeld der MRI blieb der Cobot-Arbeitsplatz durch einen Sichtschutz verdeckt. Abbildung 4.24 zeigt den Versuchsraum sowie den experimentellen Aufbau. Tabelle 31 schlüsselt die einzelnen Schritte des Versuchsablaufs detailliert auf.

¹⁰ Bei allen Items wurden die Termini in den eckigen Klammern je nach Versuchsbedingung durch die passende Bezeichnung (Paul/UR-5) ersetzt.

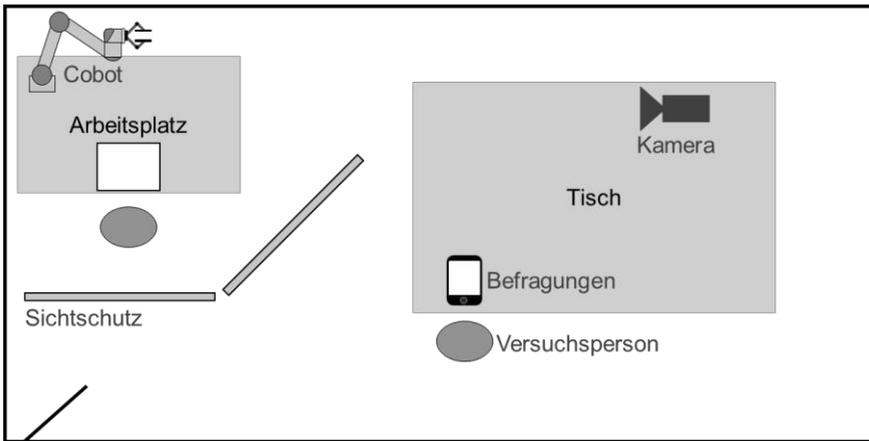


Abbildung 4.24: Skizze des Versuchsraums.

Tabelle 31: Tabellarische Darstellung des Ablaufs von Experimentalstudie 2.

Nr.	Schritt	Beschreibung
1	Einführung zur Studie	Einverständniserklärung für Videoaufnahme, Information zu anstehender Interaktion mit einem Cobot (noch hinter Sichtschutz).
2	Allgemeine Fragen	Mensch-Cobot-Relation, NARS.
3	Prä-Framing Vertrauensmessung (TAS)	Die Items bezogen sich auf einen konkreten Cobot und wurden in randomisierter Reihenfolge präsentiert.
4	Zusätzliche Fragen (Demografie, relevante Grundeinstellungen)	Geschlecht, Alter, höchster Bildungsabschluss, Studiengang, Intensität der Beschäftigung mit Robotern, vorherige Arbeit mit Robotern (ja/nein) und Kurzskala zur Kontrollüberzeugung in Bezug auf Technik von Neyer et al. (2012, S. 88).
5	Randomisierte Gruppenzuweisung	<i>Siehe Versuchsgruppen in Kapitel 4.6.1.</i>
6	Präsentation des Framing-Texts	Die Teilnehmenden wurden explizit aufgefordert, den Text gründlich zu lesen.
7	Verständnisfragen	Multiple Choice-Kontrollfrage (1 aus 5) zum Textverständnis.
8	Manipulationskontrolle (<i>manipulation checks</i>)	<i>Siehe verwendete Messmethoden zur Manipulationskontrolle in Kapitel 4.6.2.3.</i> Die Items bezogen sich konkret auf den Cobot, mit dem die Interaktion stattfand.

Nr.	Schritt	Beschreibung
9	Post-Framing Vertrauensmessung (TAS)	Die Items bezogen sich konkret auf den Cobot, mit dem die Interaktion stattfand, und wurden in randomisierter Reihenfolge präsentiert.
10	Einweisung in Interaktion	Enthüllung des Cobots, Erklärung der anstehenden Interaktion anhand bebildeter Schritt-für-Schritt-Montageanleitung.
11	Durchführung der Interaktion	<i>Siehe ausführliche Erklärung nach der Tabelle.</i>
12	Post-Interaktion Vertrauensmessung (TAS)	Die Items bezogen sich konkret auf den Cobot, mit dem die Interaktion stattfand, und wurden in randomisierter Reihenfolge präsentiert.
13	Bewertung der Interaktion und Abfrage von Akzeptanz-Items	Bewertung des Wohlbefindens während der Interaktion, der Intensität der Zusammenarbeit und ausgewählter Akzeptanz-Items.
14	Freitext-Beschreibung des Cobots	Mindestens 40 Zeichen.
15	Generelles Feedback als Freitext	Optional.
16	Dank für Teilnahme / <i>debriefing</i> / Aufwandsentschädigung	Aufwandsentschädigung in Form eines Einkaufsgutscheins der Stadt Karlsruhe im Wert von 10 Euro.

Die reale Interaktion (Schritt 11) bestand in Anlehnung an Müller-Abdelrazeq et al. (2019) darin, dass die Versuchsteilnehmenden gemeinsam mit dem fest auf einer T-Nuten-Tischplatte verschraubten Cobot UR-5 Zahnradgetriebe aus 3D-gedruckten Kunststoffteilen, Schrauben, Muttern und Gummis zusammenbauten. Dieser Zusammenbau erfolgte dreimal hintereinander, sodass insgesamt drei Zahnradgetriebe entstanden. Die Funktionsweise des Getriebes basierte dabei auf einem geringfügig angepassten CAD-Modell der *Math Gears*, das zur nicht-kommerziellen Nutzung auf der Open-Source-Plattform *Thingiverse*¹¹ zur Verfügung steht (Solo, 2013). In jedem Getriebe wurden drei unterschiedlich große Zahnräder verbaut. Bei den drei Durchgängen mussten die verschiedenen Zahnräder jeweils an unterschiedlichen Positionen innerhalb des Getriebes verbaut werden, um unterschiedliche Produktvarianten zu simulieren. Die Reihenfolge, in der die Versuchsteilnehmenden die

¹¹ Website: <https://www.thingiverse.com/>, zuletzt geprüft: 28.02.2020.

Zahnräder verbauen sollten, ergab sich aus der Bereitstellung und Positionierung durch den Cobot auf dem Arbeitstisch. Ferner befanden sich auf dem Tisch Aufbewahrungsbehälter (u. a. für Schrauben, Muttern, Gummis und Anbauteile), Haltevorrichtungen (für Zahnräder, Schrauben, Deckel und Zwischenablagen), eine Kiste für Fertigteile sowie verschiedene Kennzeichnungen (Arbeitsbereich, Zahnradablage, Lagerplätze), wie in Abbildung 4.25 zu sehen. Die Versuchspersonen erhielten einen Inbusschlüssel und eine Konstruktionsanleitung (vgl. Anhang C.X).

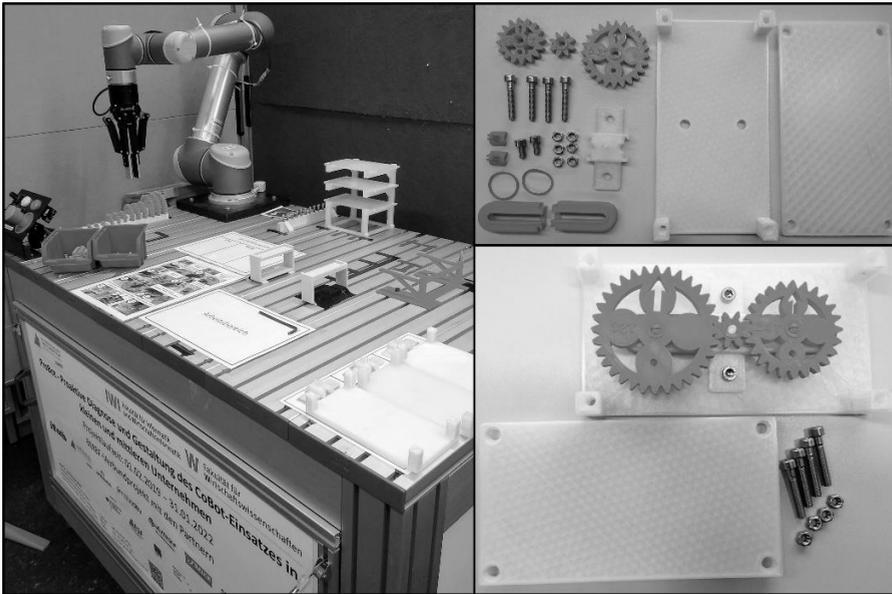


Abbildung 4.25: Arbeitsplatz und Bauteile zur Konstruktion des Getriebes.

Der genaue zeitliche Ablauf der Interaktion ist in Abbildung 4.26 visualisiert. Die MRI bestand aus kooperativen und kollaborativen Elementen. In einer ersten kooperativen Phase des Experiments übernahm der Cobot klassische Teilaufgaben der manuellen Montage, nämlich die Teilezuführung/-bereitstellung (in sinnhafter Reihenfolge) sowie die Vormontage eines Gruppenbauteils, indem der spätere Deckel des Getriebeteils mit den entsprechenden Schrauben bestückt und bereitgestellt wurde. Danach ging der Cobot in einen Wartemodus über. Die Versuchsteilnehmenden erledigten zeitgleich Montagearbeiten, die feinmotorische Fähigkeiten erfordern (Schritte 1 bis 5 in Abbildung 4.26). Während der Cobot für die einzelnen Arbeitsschritte bei allen drei Durchgängen ungefähr gleich viel Zeit benötigte, unterschied sich die Dauer für die dem menschlichen Gegenüber zugeleiteten Arbeitsschritte je nach Versuchsteilnehmenden und Durchgang. Die Dauer der Kooperation betrug beim ersten Bauteil 130 Sekunden und bei den folgenden Durchläufen

durch den notwendigen Abtransport des vorherigen Bauteils 160 Sekunden. Sobald die Versuchsteilnehmenden ihre Arbeiten erledigt hatten, reaktivierten sie den Cobot (Schritt 6) und starteten somit den zweiten Teil der Interaktion (Schritt 7 und 8).

In der anschließenden kollaborierenden Phase übernahm der Cobot die Funktion einer „dritte[n] Hand“ (Haag, 2015, S. 62), indem er das Bauteil für die Versuchsperson etwa auf Brusthöhe hielt, sodass diese beide Hände für das Befestigen der Mutter und das gleichzeitige Festziehen der Schrauben verwenden konnte. Bei dieser Kollaboration bewegte sich der UR-5 in einen gemeinsamen Arbeitsraum im vorderen Bereich des Tisches hinein. Nachdem die Versuchsperson alles verschraubt hatte, konnte sie mit einem weiteren Knopfdruck den Cobot zur Teileabführung des fertigen Werkstücks veranlassen und bereits mit der Fertigung des nächsten Getriebes beginnen.

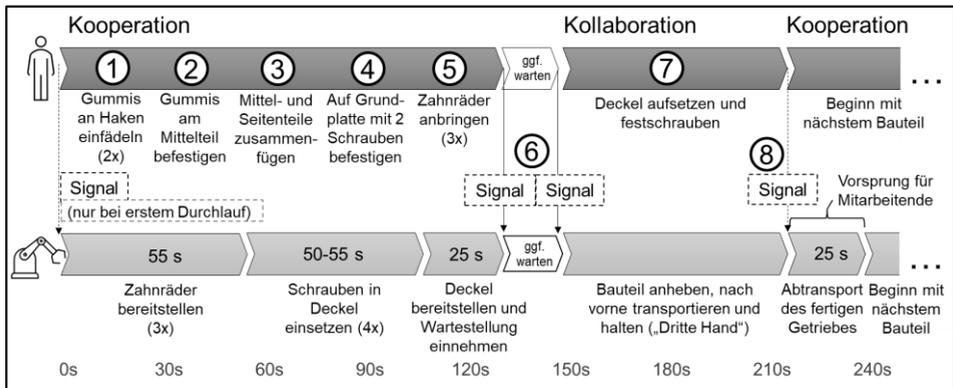


Abbildung 4.26: Zeitlicher Ablauf der Interaktion. Umkreiste Nummern repräsentieren Arbeitsschritte auf der Konstruktionsanleitung.

Laut Bøgh et al. (2012) handelt es sich bei den Tätigkeiten, die der Cobot in der experimentellen Interaktion übernahm, um Teiltätigkeiten, für die kurz- bis mittelfristig Cobots in der Praxis eingesetzt werden können. Insofern wurde der Anforderung Rechnung getragen, im Sinne einer hohen externen Validität ein möglichst realitätsnahes und somit auch möglichst wirtschaftlich sinnvolles Szenario zu konstruieren, das auch für Laien in der Komplexität bewältigbar ist und kollaborierende Elemente enthält. Das Anwendungsszenario wurde einer Risikobeurteilung nach DIN EN ISO 12100 und ISO DIN ISO/TR 14121-2 unter Berücksichtigung der Normen DIN EN ISO 10218, DIN EN ISO 13849-1 und DIN ISO/TS 15066 unterzogen. Es wurde ein akzeptables Restrisiko ermittelt, sodass keine Sicherheitsbedenken für die Zusammenarbeit vorlagen.

Die Interaktion dauerte insgesamt ca. 15 Minuten. Die Versuchsleitung beobachtete das Geschehen passiv aus einiger Entfernung. Die Durchführung des kompletten Experiments

dauerte pro Versuchsperson ca. 30 bis 45 Minuten. Das Experiment wurde als doppelblinde Studie durchgeführt, d. h. weder dem Experimentator noch der Versuchsperson war die jeweilige Zuordnung zur Versuchsbedingung bekannt (vgl. Wullenkord & Eyszel 2019).

4.6.2.5 Statistische Analysen

Grundsätzlich ähneln sich die statistischen Analysen von Experimentalstudie 1 und Experimentalstudie 2 stark (vgl. Kapitel 4.5.2.5). Da die wahrgenommene Kooperationsorientierung im Gegensatz zu Experimentalstudie 1 nicht erhoben wurde und somit nur ein Zwischensubjektfaktor vorlag, war in Experimentalstudie 2 zur Auswertung der Effekte auf das Vertrauen allerdings eine zweifaktorielle gemischte 2×3 -ANOVA mit dem zweistufigen Zwischensubjektfaktor der wahrgenommenen Menschenähnlichkeit, dem Messzeitpunkt als dreistufigem Innersubjektfaktor und jeweils den absoluten Vertrauens- und Misstrauenswerten als AV vorgesehen. Deren Anwendung erübrigte sich allerdings im weiteren Verlauf der Auswertungen aufgrund einer unwirksamen experimentellen Manipulation. Um dennoch zumindest explorativ evaluieren zu können, inwiefern die Bewertung der durchgeführten Interaktion vom empfundenen Vertrauen abhängt, wurden die Versuchsteilnehmenden mittels eines sog. *median splits* entweder einer Gruppe mit unter- oder mit überdurchschnittlichem Vertrauen zugeordnet. Die Gruppenvergleiche wurden anhand von t -Tests durchgeführt (vgl. Kapitel 4.6.3.5). Analog zu Experimentalstudie 1 kamen bei der Manipulationskontrolle ebenfalls t -Tests und Mann-Whitney- U -Tests für Gruppenvergleiche zum Einsatz. Zur Analyse der Vertrauens- und Misstrauensentwicklung über die drei Messzeitpunkte hinweg wurden Wilcoxon-Tests eingesetzt. Für die Auswahl der Testverfahren und die Darstellung der Ergebnisse gelten die Erläuterungen aus Experimentalstudie 1 (siehe Kapitel 4.5.2.5).

4.6.3 Ergebnisse

4.6.3.1 Manipulationskontrolle

Ein zweiseitiger unabhängiger t -Test ergab keine signifikanten Unterschiede in der wahrgenommenen Menschenähnlichkeit zwischen den beiden Versuchsgruppen, $t(66) = -.371$, $p = .712$. Wie in Abbildung 4.27 ersichtlich, wurde der menschenähnlich beschriebene Cobot erwartungskonform als menschenähnlicher wahrgenommen. Allerdings unterschieden sich die Mittelwerte pro Versuchsgruppe im Gegensatz zu Experimentalstudie 1 kaum voneinander.

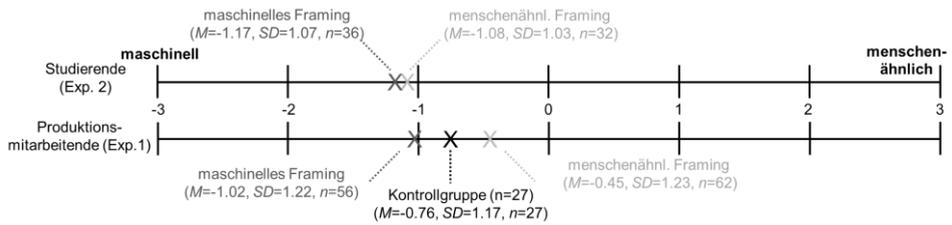


Abbildung 4.27: Grafische Darstellung der wahrgenommenen Menschenähnlichkeit (*Godspeed*-Skala) je Versuchsgruppe mit Vergleich zu den Ergebnissen aus Experimentalstudie 1.

Aufgrund dieses überraschenden Ergebnisses wurden die Werte für die wahrgenommene Menschenähnlichkeit näher untersucht. In einem ersten Ansatz wurden dazu nur diejenigen Versuchsteilnehmenden in die Berechnung einbezogen, die analog zu Experimentalstudie 1 über keine praktischen Vorerfahrungen verfügten. Ein weiterer Ansatz bestand darin, die ausgeschlossenen zwölf Datensätze einzubeziehen, bei denen es zu einer fehlerhaften Interaktion gekommen war, um die Teststärke zu erhöhen. Wie Tabelle 32 zeigt, ergaben sich allerdings auch für diese Stichprobenszusammensetzungen keine signifikanten Unterschiede. Das gleiche Bild ergab sich bei einer Unterteilung der Stichprobe anhand der Studiengänge in eher technisch bzw. nicht-technisch orientierte Studienfächer (vgl. Anhang C.IX). In Anhang VI ist die Verteilung der Daten pro Versuchsgruppe zusätzlich als Box-Plot dargestellt.

Tabelle 32: Ergebnisse der Signifikanztests auf Unterschiede in der wahrgenommenen Menschenähnlichkeit zwischen den Versuchsgruppen, je nach Stichprobenszusammensetzung.

	<i>n</i>	Menschen- ähnl. Framing		Maschi- nelles Framing		<i>t</i> -Test		
		<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>df</i>	<i>t</i>	<i>p</i>
Reguläre Stichprobe	68	-1.08	1.03	-1.17	1.01	66	-.371	.712
Ohne praktische Vorerfahrungen	59	-1.04	0.98	-1.15	0.87	47	.415	.680
Mit fehlerhafter Interaktion	80	-1.11	1.03	-1.22	0.99	78	-.509	.612

Analog zu Experimentalstudie 1 wurde die wahrgenommene Menschenähnlichkeit zusätzlich anhand einer Bewertung der Freitext-Beschreibungen des Cobots durch die Versuchsteilnehmenden auf einer fünfstufigen Skala von 1 = *maschinell* bis 5 = *menschähnlich* abgebildet. Abweichend zu Experimentalstudie 1 verfassten die Versuchsteilnehmenden die Freitexte erst nach der Durchführung der MRI, was deren Aussagekraft aufgrund des möglichen Einflusses weiterer Variablen wie der Wahrnehmung der MRI verringert. Der

Mann-Whitney- U -Test als non-parametrische Alternative zum unabhängigen t -Test zeigte keine signifikanten Unterschiede zwischen den beiden Versuchsgruppen, $U = 1225.00$, $Z = -0.227$, $p = .820$. Abbildung 4.28 verdeutlicht, dass sich die Mittelwerte pro Versuchsgruppe kaum unterscheiden.

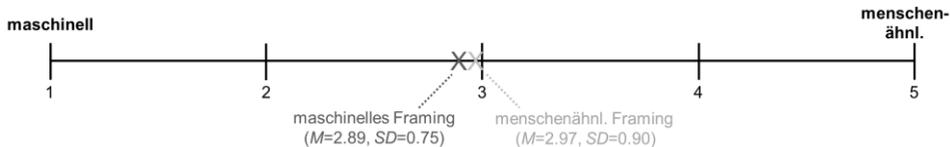


Abbildung 4.28: Grafische Darstellung der wahrgenommenen Menschenähnlichkeit (Bewertung gemäß Kordierleitfaden als indirekte Messmethode) je Versuchsgruppe.

Beide Auswertungen legen nahe, dass die experimentelle Manipulation keinen signifikanten Effekt hervorrief. Damit erübrigten sich die geplanten varianzanalytischen Analysen zur Auswirkung der unterschiedlichen Versuchsbedingungen auf das Vertrauensniveau. Stattdessen fokussieren die folgenden Auswertungen auf die Vertrauens- und Misstrauensentwicklung über die Zeit. Dargestellte Unterschiede zwischen den beiden Versuchsgruppen sind als explorative Analyse anzusehen.

4.6.3.2 Vertrauensentwicklung

Abbildung 4.29 zeigt den Verlauf des Vertrauens in den Cobot über die drei Messzeitpunkte hinweg. Wie Wilcoxon-Tests nahelegen, änderten sich die Vertrauenswerte weder durch das menschenähnliche, $Z = -0.393$, $p = .695$, noch durch das maschinelle Framing signifikant, $Z = -1.229$, $p = .219$. Allerdings lagen die Vertrauenswerte nach der Interaktion sowohl in der menschenähnlichen, $Z = -4.337$, $p < .001$, als auch in der maschinellen Bedingung, $Z = -4.224$, $p < .001$, signifikant höher als vor der Interaktion.

Da sich Fehler in der Interaktion mutmaßlich besonders stark auf die Vertrauensentwicklung auswirken, wurden die zwölf Versuchsdurchläufe mit unbeabsichtigt zustande gekommenen Fehlern in der MRI zu einer zusätzlichen Vergleichsgruppe zusammengefasst und ebenfalls in Abbildung 4.29 dargestellt. Es zeigt sich insbesondere eine weniger starke, nicht-signifikante Steigerung des Vertrauensniveaus durch die fehlerhafte MRI, $Z = -1.721$, $p = .085$. Bei der interferenzstatistischen Analyse ist allerdings zu beachten, dass die Teststärke aufgrund der wesentlich kleineren Stichprobe geringer ausfällt und damit die Wahrscheinlichkeit eines Fehlers zweiter Art höher liegt.

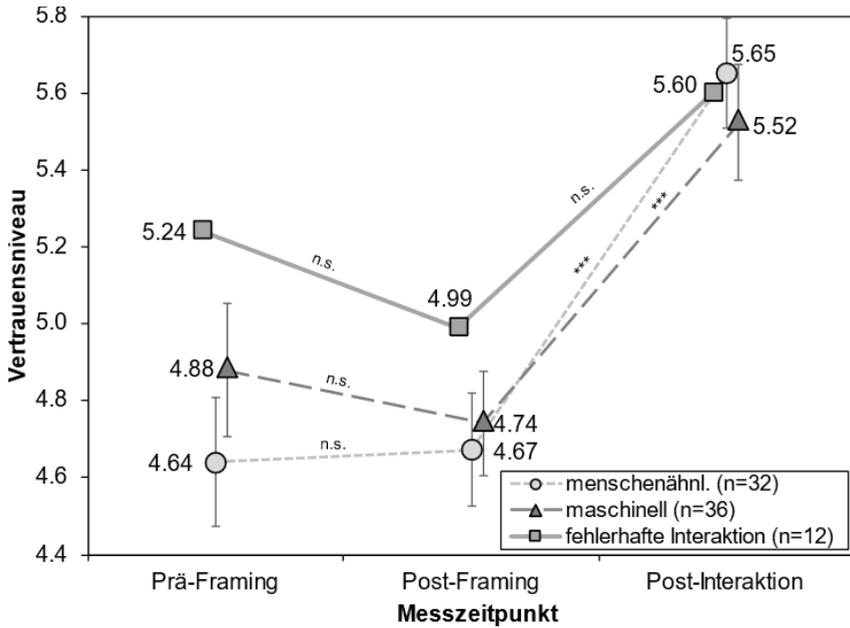


Abbildung 4.29: Grafische Darstellung des gemessenen Vertrauensniveaus an allen drei Messzeitpunkten je Versuchsgruppe und im Vergleich zu den Versuchsteilnehmenden, bei denen Fehler in der Interaktion auftraten. Aus Gründen der Übersichtlichkeit sind bei letztgenannter Gruppe keine Fehlerbalken eingeblendet.

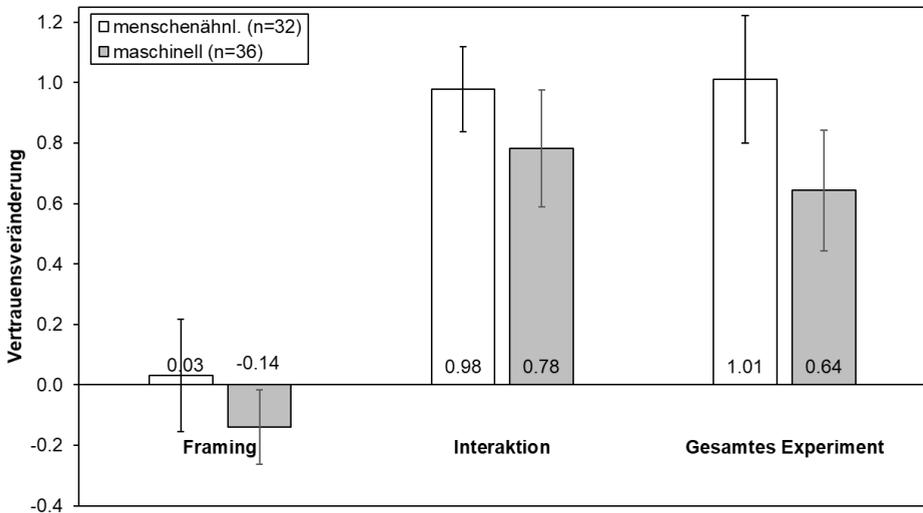


Abbildung 4.30: Grafische Darstellung der Vertrauensveränderung zwischen den Messzeitpunkten je Versuchsgruppe. Alle Gruppenunterschiede n. s. ($ps > .065$).

Abbildung 4.30 zeigt die Vertrauensveränderung je Versuchsgruppe zwischen den verschiedenen Messzeitpunkten. In allen Vergleichszeiträumen zeigte sich deskriptiv ein stärkerer Vertrauenszuwachs in derjenigen Gruppe, die das menschenähnliche Framing erhielt. Laut Mann-Whitney- U -Tests handelt es sich hierbei allerdings jeweils um keine signifikanten Unterschiede zwischen den Versuchsgruppen, wie in Tabelle 33 ersichtlich.

Tabelle 33: Übersicht über gruppenspezifisch unterschiedliche Vertrauensveränderungen durch verschiedene Interventionen.

	Mann-Whitney- U -Test		
	U	Z	p
Framing	496.00	-0.986	.324
Interaktion	530.00	-0.567	.571
Gesamtes Experiment (Framing & Interaktion)	426.00	-1.848	.065

4.6.3.3 Misstrauensentwicklung

Analog zur Vertrauensentwicklung zeigt Abbildung 4.31 den Verlauf des Misstrauens in den Cobot über die drei Messzeitpunkte hinweg.

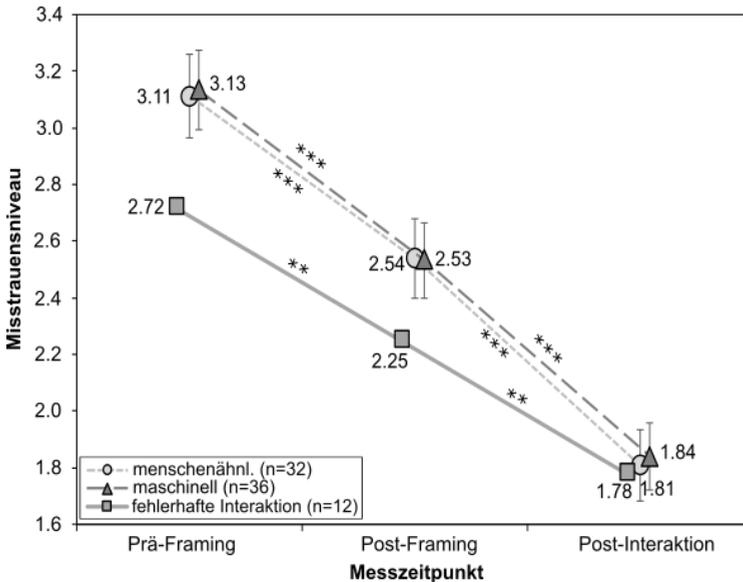


Abbildung 4.31: Grafische Darstellung des gemessenen Misstrauensniveaus an allen drei Messzeitpunkten je Versuchsgruppe und im Vergleich zu den Versuchsteilnehmenden, bei denen Fehler in der Interaktion auftraten. Aus Gründen der Übersichtlichkeit sind bei letztgenannter Gruppe keine Fehlerbalken eingeblendet.

Wie Wilcoxon-Tests zeigen, änderte sich das empfundene Misstrauen in beiden Versuchsbedingungen zwischen allen Messzeitpunkten signifikant ($ps < .001$). Im Gegensatz zur Vertrauensänderung wirkte sich demnach sowohl das Framing als auch die Durchführung der MRI signifikant auf das empfundene Misstrauen aus. In der Versuchsgruppe, bei der es zu Fehlern in der Interaktion kam, lassen sich zwar deskriptiv geringere, aber ebenfalls signifikante Veränderungen zwischen den Messzeitpunkten feststellen, $Z_{\text{Framing}} = -2.729$, $p_{\text{Framing}} = .006$, $Z_{\text{Interaktion}} = -2.669$, $p_{\text{Interaktion}} = .008$.

Wie sich in Abbildung 4.32 unschwer deskriptiv erkennen lässt, sind zu keinem Messzeitpunkt signifikante Unterschiede zwischen den Versuchsgruppen in Hinblick auf die Misstrauensabnahme feststellbar ($ps > .805$).

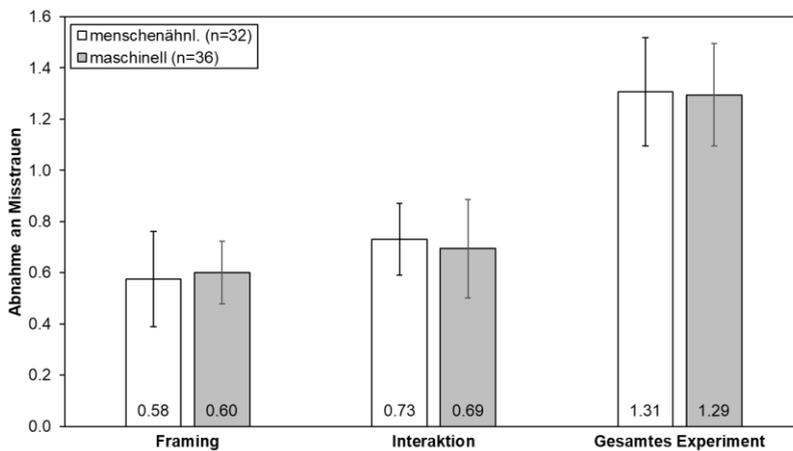


Abbildung 4.32: Grafische Darstellung der Abnahme an Misstrauen zwischen den Messzeitpunkten je Versuchsgruppe. Alle Gruppenunterschiede n. s. ($ps > .805$).

4.6.3.4 Auswertung der Akzeptanz-Items

Abbildung 4.33 stellt die Zustimmungswerte zu den nach der Interaktion abgefragten Akzeptanz-Items dar (vgl. Kapitel 4.6.2.3). Da laut Mann-Whitney- U -Tests keine signifikanten Unterschiede zwischen den Versuchsgruppen vorlagen ($ps > .078$), differenziert diese explorative Auswertung nicht nach Versuchsgruppen.

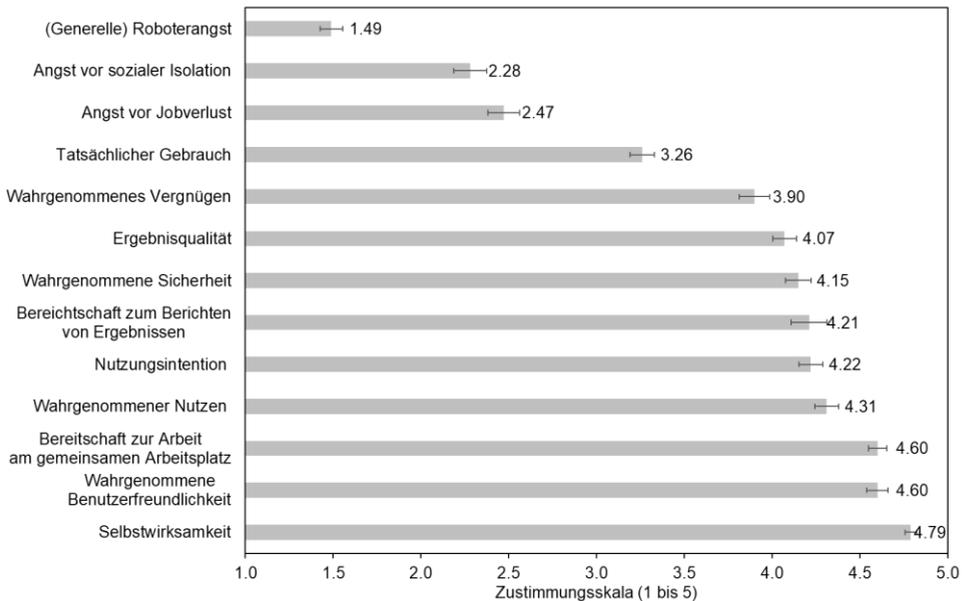


Abbildung 4.33: Deskriptivstatistische Auswertung der Zustimmungswerte zu den abgefragten Akzeptanz-Items.

Die höchste Zustimmung erzielten Aussagen, die sich auf die Sicherheit im Umgang mit dem Roboter sowie dessen Benutzungsfreundlichkeit und Nutzen beziehen. So trauten sich die Versuchsteilnehmenden den Umgang mit einem Cobot zu (Selbstwirksamkeit; $M = 4.79$, $SD = 0.41$), betrachteten dessen Nutzung als einfach (wahrgenommene Benutzerfreundlichkeit; $M = 4.60$, $SD = 0.74$) und zeigten eine hohe Bereitschaft, an einem gemeinsamen Arbeitsplatz mit diesem zu arbeiten (Bereitschaft zur Arbeit am gemeinsamen Arbeitsplatz; $M = 4.60$, $SD = 0.63$). Außerdem wurde die Nützlichkeit als hoch bewertet (wahrgenommener Nutzen; $M = 4.31$, $SD = 0.82$). Verschiedene Ängste in Bezug auf Cobots sind hingegen nur wenig verbreitet. So würden sich die Versuchsteilnehmenden im Schnitt nicht unwohl in der Gegenwart des Cobots fühlen (Roboterangst; $M = 1.49$, $SD = 0.76$) und hätten auch nur wenig Angst vor sozialer Isolation ($M = 2.28$, $SD = 1.12$) und vor einem Jobverlust ($M = 2.47$, $SD = 1.09$).

4.6.3.5 Bewertung der Interaktion

In einer weiteren explorativen Analyse wurden die Versuchsteilnehmenden mittels eines *median splits* anhand ihres Vertrauenswertes vor der Interaktion entweder in eine Gruppe mit unterdurchschnittlichem Vertrauen ($M = 4.00$, $n = 32$) oder mit überdurch-

schnittlichem Vertrauen ($M = 5.39, n = 33$) eingeteilt.¹² Anschließend wurden die Bewertungen der Interaktion zwischen diesen Gruppen verglichen, wie in Abbildung 4.34 ersichtlich.

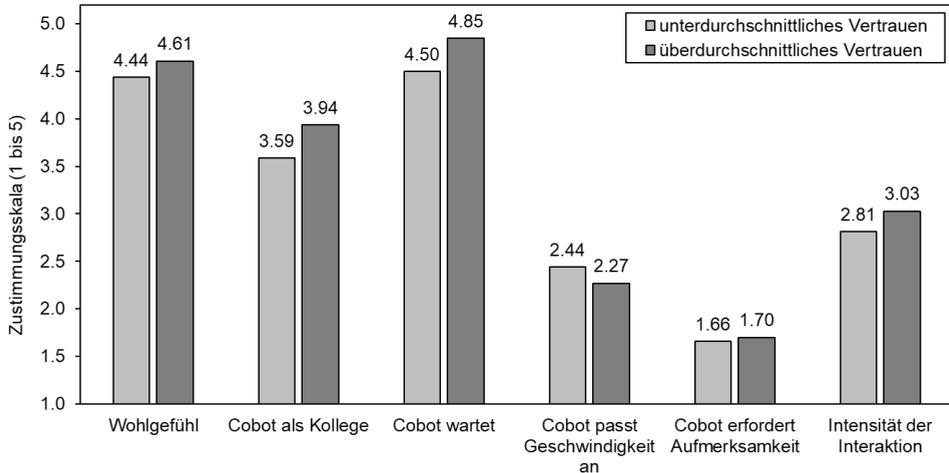


Abbildung 4.34: Grafische Darstellung der Zustimmungswerte zu Aussagen bezogen auf das Erleben der Interaktion, differenziert nach Untermenge der Stichprobe mit unter- bzw. mit überdurchschnittlichem Vertrauen.

Zwar sind keine Unterschiede signifikant ($ps > .091$), dennoch zeigt sich – mit Ausnahme der angepassten Arbeitsgeschwindigkeit – ein deskriptiver Trend dergestalt, dass die Versuchsteilnehmenden mit überdurchschnittlichem Vertrauen die Interaktion positiver wahrnahmen als diejenigen mit unterdurchschnittlichem Vertrauen. Diese Tendenz wird zusätzlich dadurch gestützt, dass sich die Versuchspersonen mit unterdurchschnittlichem Misstrauen einer analog durchgeführten Auswertung zufolge während der Interaktion signifikanter wohler fühlten als die Versuchspersonen mit überdurchschnittlichen Misstrauen, $t(55) = 3.180, p = .002$.

Eine Betrachtung der Mittelwerte über beide Versuchsgruppen hinweg zeigte ferner, dass sich die Versuchsteilnehmenden während der Interaktion sehr wohl fühlten ($M = 4.51, SD = 0.68$), den Cobot überwiegend als Kollegen empfanden ($M = 3.78, SD = 1.09$), das Gefühl hatten, dass der Cobot auf sie wartet ($M = 4.69, SD = 0.82$) und ihre Aufmerksamkeit nur in geringem Maße beansprucht sahen ($M = 1.72, SD = 0.88$). Selbst in diesem

¹² Drei Versuchsteilnehmende lagen direkt auf dem Median und wurden von der Analyse ausgenommen, um Verzerrungen der Gruppen zu vermeiden. Bei der analogen Analyse in Bezug auf das Misstrauen betraf dies elf Versuchsteilnehmende.

Szenario mit echt kollaborierenden Tätigkeitsanteilen wurde die Intensität der Interaktion nur als mittelmäßig bewertet ($M = 2.94$, $SD = 0.84$).

Eine vorgesehene Auswertung des Verhaltens auf Basis der Videoaufnahmen lieferte keine zusätzlichen Erkenntnisse, da die *inter-rater*-Reliabilität bei der Kodierung des Videomaterials zu gering ausfiel.

4.6.3.6 Explorative Korrelationsanalyse

Zwischen dem neu eingeführten Multi-Item-Messinstrument zur Technikkontrollüberzeugung (vgl. Kapitel 4.6.2.3) und dem auch in Experimentalstudie 1 verwendeten Single-Item-Messinstrument zeigte sich erwartungsgemäß ein signifikanter stark positiver Zusammenhang, $r = .538$, $p < .001$. Das spricht für eine ausreichend hohe Reliabilität des Single-Item-Messinstruments. Bei den weiteren Auswertungen wird aufgrund der Überlegenheit validierter Multi-Item-Skalen ausschließlich auf das Multi-Item-Konstrukt zurückgegriffen.

Während Vertrauen und Misstrauen sowohl vor dem Framing, $r = -.497$, $p < .001$, als auch nach dem Framing, $r = -.345$, $p = .004$, signifikant mittelmäßig bis stark negativ korrelierten, zeigte sich nach der Interaktion kein signifikanter Zusammenhang mehr, $r = -.138$, $p = .261$, was auf eine unterschiedliche Beeinflussung der beiden Konstrukte durch die Interaktion hindeutet. Ferner lässt sich auf Basis der Korrelationen konstatieren, dass eine höhere wahrgenommene Menschenähnlichkeit in Zusammenhang mit einer positiveren Grundeinstellung gegenüber Robotern steht, $r = -.301$, $p = .013$. Ein hohes Vertrauensniveau nach der Interaktion geht einher mit der Zustimmung zu der Aussage, dass sich die Versuchsteilnehmenden während der Interaktion wohlgefühlt haben, $r = .320$, $p = .008$. Das Wohlgefühl während der Interaktion weist damit unter allen Items, die zur Bewertung der Interaktion erhoben wurden, den stärksten Zusammenhang zum Vertrauen auf.

Überdies scheint die Grundeinstellung gegenüber Robotern (NARS) ein recht stabiler Prädiktor für das Vertrauen in einen Cobot zu sein. Es zeigten sich signifikante Korrelationen sowohl zum initialen Vertrauen vor dem Framing, $r = -.283$, $p = .019$, als auch zum Vertrauen nach dem Framing, $r = -.328$, $p = .006$, und nach der Interaktion, $r = -.323$, $p = .007$. Ferner bestehen analog zu Experimentalstudie 1 signifikante Zusammenhänge zwischen der Grundeinstellung und dem Vorwissen, $r = -.262$, $p = .031$, bzw. zwischen der Grundeinstellung und der Technikkompetenzüberzeugung, $r = -.280$, $p = .021$. Tabelle 34 stellt außerdem die Korrelationen zwischen dem Vertrauen bzw. dem Misstrauen nach der Interaktion und den jeweiligen Akzeptanz-Items dar, um den Zusammenhang zwischen diesen beiden oft synonym verwendeten Konstrukten näher zu explorieren.

Tabelle 34: Signifikante Korrelationen zwischen den Akzeptanz-Items und dem gemessenen Vertrauen sowie dem gemessenen Misstrauen, jeweils nach der Interaktion.

	Vertrauen nach der Interaktion		Misstrauen nach der Interaktion	
	<i>r</i>	<i>p</i>	<i>r</i>	<i>p</i>
Wahrgenommenes Vergnügen	.268	.027*	-.110	.371
Bereitschaft zur Arbeit am gemeinsamen Arbeitsplatz	.394	.001**	-.241	.048*
Wahrgenommene Sicherheit	.514	< .001***	-.273	.025*
Wahrgenommener Nutzen	.267	.028*	.008	.949
Nutzungsintention	.284	.019*	-.188	.126
Angst vor sozialer Isolation	.197	.145	.338	.005*
Angst vor Jobverlust	-.119	.104	.389	< .001***
(Generelle) Roboterangst	-.059	.631	.448	< .001***
Selbstwirksamkeit	-.107	.383	-.066	.595
Wahrgenommene Benutzerfreundlichkeit	.157	.202	-.142	.248
Bereitschaft zum Berichten der Ergebnisse	.089	.469	.066	.594
Ergebnisqualität	.170	.166	-.173	.157
Tatsächlicher Gebrauch	.195	.110	-.067	.586

4.6.3.7 Qualitative Analyse der Freitext-Aussagen

In die qualitative Analyse flossen sowohl die Freitext-Beschreibungen des Cobots, die bereits für die Manipulationskontrolle kodiert wurden, als auch die freiwilligen Freitext-Kommentare am Ende der Umfrage ein. In beiden Versuchsgruppen zeigte sich, dass der Cobot überwiegend als nützliche (sieben Nennungen), hilfreiche und zuverlässige (17 Nennungen) Erleichterung und Unterstützung (15 Nennungen) bei der Arbeit aufgefasst wurde:

- „Paul ist ein Roboter, der mich bei meiner Arbeit unterstützt“ (70, *Paul*)

In den Kommentaren der Zugehörigen beider Versuchsgruppen fanden sich vorwiegend technische Umschreibungen des Cobots, wie die folgenden Zitate illustrieren:

- „Der UR-5 ist ein kleiner 5-achsiger Robotergreifarm, der sich mithilfe weniger Knöpfe bedienen lässt“ (5, *UR-5*)

- „Ein Roboterarm mit 2 oder 3 Drehachsen, der sich langsam und sehr geräuscharm bewegt und simple Arbeitsschritte übernimmt“ (39, *Paul*).

Vereinzelt beschrieben die Versuchsteilnehmenden den Cobot mit stark vermenschlichen Adjektiven, wie z. B. „niedlich“ (43, *UR-5*), „hilfsbereit“ (60, *UR-5*), „knuddelig“ (79, *UR-5*), „nett“ und „sympathisch“ (jeweils 82, *Paul*). Einige Versuchsteilnehmende betonten ferner, dass ihnen die Interaktion Spaß gemacht habe, zumal die Durchführung der Aufgabe mit einem Cobot „einfach mehr Spaß als alleine“ (29, *UR-5*; 13, *Paul*) mache. Eine Versuchsperson bewertete positiv, dass der Cobot „einem nicht zu viel Arbeit weg[nimmt], sodass man auch noch selbst etwas arbeiten kann und nicht die gesamte Arbeit von einem Roboter übernommen werden würde“ (26, *Paul*). Eine andere Versuchsperson äußerte Verständnis dafür, „wieso sich manch einer ‚bedroht‘ fühlen könnte, da die Weiterentwicklung des Roboters sein könnte, das gesamte Bauteil alleine zu produzieren“ (84, *UR-5*).

Außerdem lieferten die Aussage der Versuchsteilnehmenden Hinweise darauf, dass eine anfängliche Skepsis im Umgang mit dem Cobot schnell einem angenehmen Gefühl der Vertrautheit und Sicherheit wich:

- „Ich hatte am Anfang ein komisches Gefühl mit einem Roboter zu arbeiten, aber es war sehr angenehm“ (56, *Paul*)
- „Ich [war] schon nach den 3 Bauteilen mit Paul (zumindest bei diesem Arbeitsablauf) vertraut“ (14, *Paul*)
- „Durch die Arbeit mit ihm verliert man schnell die Angst oder die Vorsicht vor ihm, was sehr positiv ist“ (79, *UR-5*)

Viele Studienteilnehmende erwähnten überdies die zu niedrige und unangepasste Arbeitsgeschwindigkeit des Cobots. Manche Befragten fühlten sich allerdings auch gestresst vom schnell arbeitenden Roboter:

- „Am Anfang war er schneller fertig als ich; das hat mich etwas gestresst, ein bisschen wie bei einem Wettlauf“ (46, *UR-5*)

Zwei Teilnehmende in der Versuchsbedingung mit dem menschenähnlichen Framing äußerten sich explizit kritisch in Bezug auf die vermenschlichende Beschreibung, die sie offenbar als unpassend empfanden und damit einen möglichen Erklärungsansatz für die unwirksame experimentelle Manipulation lieferten:

- „Ich finde die Vermenschlichung irgendwie komisch. Es ist nun mal ein Roboter und nicht *Paul* :D“ (58, *Paul*)

- „Dennoch finde ich die Persona, die ihm zugewiesen wurde, merkwürdig. Ich hätte ihm einen weniger menschlichen Namen verliehen, der mehr zu einem Roboterarm passt“ (34, Paul)

4.6.4 Diskussion

4.6.4.1 Mangelnde Wirksamkeit des Framings

Entgegen der Erwartung zeigte die experimentelle Manipulation in Experimentalstudie 2 keine signifikante Wirkung, weswegen sich auf Basis der vorliegenden Daten auch keine Erkenntnisse in Hinblick auf die Persistenz von Framing-Effekten ableiten lassen (FF5). Zwar bewerteten die Versuchspersonen die wahrgenommene Menschenähnlichkeit des Cobots geringfügig höher, wenn sie ein menschenähnliches anstatt eines maschinellen Framings erhielten, allerdings fiel diese Diskrepanz wesentlich geringer aus als in Experimentalstudie 1 und gilt nicht als signifikant. Die Bewertungen anhand der *Godspeed*-Skala und anhand der Analyse der Freitext-Beschreibungen ergaben ein übereinstimmendes Gesamtbild. Die Freitext-Beschreibungen vermittelten überwiegend ein stark technisches Verständnis des Cobots und enthielten nur selten anthropomorphe Metaphern und Umschreibungen. An dieser Stelle schlägt sich womöglich auch ein *sampling bias* nieder, da tendenziell technik- und insbesondere Robotik-interessierte Studierende dem Aufruf zur Teilnahme am Experiment folgten, die den Cobot mehrheitlich als nützliches Werkzeug und als Unterstützung beim Arbeitsprozess empfanden.

Da in Experimentalstudie 2 annähernd die gleichen Texte wie in Experimentalstudie 1 verwendet wurden, stellt sich die Frage, wie sich die erhebliche Diskrepanz der Ergebnisse begründen lässt. Hierfür können drei Hypothesen ins Feld geführt werden. Besonders augenfällig ist erstens der Unterschied in der rekrutierten Stichprobe, die aus Studierenden im Vergleich zu Produktionsmitarbeitenden bestand. Die Stichproben werden in Kapitel 4.7.1 einem ausführlichen statistischen Vergleich unterzogen. Zweitens besteht die Möglichkeit, dass der Unterschied durch die Differenz zwischen Online- und Präsenz-Setting zustande kam. In Experimentalstudie 2 könnte das Wissen der Versuchsteilnehmenden um die anstehende reale Interaktion mit dem Cobot dazu geführt haben, dass diese den einführenden Informationen nur wenig Bedeutung beimaßen und sich gedanklich bereits auf die reale MRI fokussierten, wodurch womöglich affektive im Vergleich zu kognitiven Vertrauensanteilen stärker ins Gewicht fielen. Drittens könnte die mangelnde Nachweisbarkeit eines Effekts durch die geringere Stichprobenanzahl bedingt sein. So ergab eine *a posteriori*-Teststärken-Analyse unter Annahme der Effektstärke aus Experimentalstudie 1 ($\eta^2 = .051$) eine geringe Teststärke von $(1-\beta) = 47\%$.

Die Bemerkungen zweier Versuchsteilnehmenden innerhalb der Freitext-Kommentare nähren die erstgenannte Hypothese, dass die Effektivität des Framings – abgesehen von

den methodischen Einflüssen – stark von der Population abhängt. Diese Studierenden bezeichneten die vermenschlichenden Umschreibungen im Text als merkwürdig und distanzieren sich deutlich davon. In diesen Fällen liegt die Vermutung nahe, dass die Studierenden bereits über ein gefestigtes mentales Modell verfügten, das Roboter als Maschinen repräsentierte. Wenn im Framing präsentierte Informationen dem (vermeintlichen) Wissensstand der entsprechenden Person deutlich widersprechen und sich daher nicht ohne Weiteres in bestehende Wissensstrukturen assimilieren lassen, kann dies zu kognitiver Dissonanz führen (Festinger, 1957). Diese ist dadurch gekennzeichnet, dass eine Person inkonsistente Kognitionen erlebt, also bspw. einen Widerspruch zwischen einer Wahrnehmung von präsentierten Informationen und einer bestehenden Überzeugung. Aufgrund des „Streben[s] des Menschen nach Widerspruchsfreiheit“ (Raab, Unger & Unger, 2010, S. 42) wird diese Dissonanz als unangenehm erlebt und versucht, sie aufzulösen. Das ließe sich durch die Anpassung bestehender Wissensstrukturen realisieren, d. h. in diesem Fall durch die Änderung der rein technischen Repräsentation des Roboters, sodass diese zu den vermittelten Informationen passt. Solche Modifikationen oder Neubildungen mentaler Modelle geschehen stets als Folge einer Abwehrhaltung gegenüber einer bloßen Assimilation von Wahrnehmungen in bestehende Strukturen (Seel et al., 2009, S. 25). Im Kontext von Technologieadoptionsprozessen verfolgen Personen allerdings mitunter eher die Strategie, ihre bestehende Wissensbasis gegenüber widerstrebenden Informationen zu schützen, indem sie die Rezeption solcher Informationen vermeiden, sie umdeuten und/oder bevorzugt konsonante Informationen selektiv konsumieren (Marikyan, Papagiannidis & Alamanos, 2020). Entsprechend lässt sich die Reaktion der Versuchsteilnehmenden gegenüber den vermenschlichenden Informationen aus dem Framing als Versuch werten, ihr maschinelles mentales Modell von Robotern gegen kognitive Dissonanz zu schützen, indem sie diese Informationen als gezielten Manipulationsversuch der Versuchsleitung deklarierten. In der Gesamtschau scheint daher die Annahme plausibel, dass die Effektivität von Framing stark von den beteiligten Personen und deren existenten mentalen Modellen abhängt. Dass sich inferenzstatistisch kein signifikanter Effekt der experimentellen Manipulation nachweisen ließ, könnte in Anbetracht der *a posteriori*-Teststärken-Analyse auf ein Zusammenspiel einer populationsbedingt geringeren Effektgröße des Framings sowie einer insgesamt geringeren Teststärke aufgrund der niedrigeren Stichprobenanzahl zurückzuführen sein.

4.6.4.2 Wahrnehmung der MRI

Aufgrund der fehlgeschlagenen experimentellen Manipulation und den damit verbundenen nicht-signifikanten Unterschieden zwischen den Versuchsgruppen ließ sich die vergleichende Betrachtung der Bewertungen der Interaktion nicht wie geplant durchführen. Alternativ wurden alle Versuchsteilnehmenden mittels eines *median splits* auf Basis des Vertrauenswertes vor der Interaktion in zwei Gruppen unterteilt. Der Vergleich dieser beiden Gruppen offenbarte keine signifikanten Unterschiede. Allerdings zeichnete sich eine deskriptive Tendenz dahingehend ab, dass die Versuchsteilnehmenden mit

überdurchschnittlichem Vertrauen die Interaktion als positiver und intensiver bewerteten. Sie fühlten sich während der Interaktion wohler und nahmen den Cobot eher als Kolleg:in wahr. Insofern liegt die Vermutung nahe, dass einstellungsbasiertes initiales Vertrauen auch zu einer positiveren Wahrnehmung einer realen MRI führt und somit innerhalb einer Rückkopplungsschleife das dynamische Vertrauen positiv beeinflusst. Allerdings können diese Ergebnisse insgesamt aufgrund der quasi-experimentellen Gruppenzuordnung und der nicht-signifikanten Unterschiede nur als schwache empirische Evidenz für vertrauensbasierte Unterschiede in der Wahrnehmung einer MRI gewertet werden (*FF4*). Die vergleichsweise geringen zahlenmäßigen Unterschiede in den Bewertungen könnten auch einem erhebungsbedingtem Deckeneffekt geschuldet sein, da der Mittelwert für das Wohlfühl während der Interaktion mit $M = 4.44$ bereits bei der Versuchsgruppe mit unterdurchschnittlichem Vertrauen im oberen Bereich der fünfstufigen Skala lag.

4.6.4.3 Vertrauen und Misstrauen

In Experimentalstudie 2 zeigte sich kein vertrauenssteigernder Effekt durch die Lektüre der Framing-Texte in den Experimentalgruppen. Wenngleich die Diskussion vorwiegend auf das Konstrukt des Vertrauens rekurriert, bietet sich an dieser Stelle auch eine explizite Betrachtung des Misstrauens an, da sich dieses in Experimentalstudie 2 nicht antiproportional zum Vertrauensniveau entwickelte. Vielmehr blieb das Vertrauen zwischen den Messzeitpunkten vor und nach dem Framing gleich, wohingegen sich das Misstrauen in beiden Versuchsgruppen signifikant reduzierte. Fraglich ist in diesem Zusammenhang, ob dieser Effekt, wie in Experimentalstudie 1 vermutet, ebenfalls dem Effekt des imaginierten Kontakts zugerechnet werden kann, zumal bis dato keine Studien vorliegen, die diesen Effekt unter Differenzierung von Vertrauen und Misstrauen als Zielvariable untersuchen. Eine nähere Analyse der Unterschiede zwischen beiden Experimentalstudien findet sich in Kapitel 4.7.3.

Die bemerkenswerte Diskrepanz in den durch Framing hervorgerufenen Entwicklungen von Vertrauen und Misstrauen verweisen auf deren unterschiedliche Wirkmechanismen und Einflussfaktoren. Bei Betrachtungen der in Experimentalstudie 2 berechneten Korrelationen zwischen Vertrauen, Misstrauen und einzelnen Aspekten der Akzeptanz fällt auf, dass die Zusammenhänge der Akzeptanzkomponenten zu Vertrauen und Misstrauen je unterschiedlich ausfallen. Nur die wahrgenommene Sicherheit und die Bereitschaft zur Arbeit an einem gemeinsamen Arbeitsplatz weisen sowohl signifikante Korrelationen zu Vertrauen als auch zu Misstrauen auf, wie Abbildung 4.35 visualisiert. Misstrauen scheint stärker mit Befürchtungen verknüpft zu sein, wie z. B. der Angst vor Jobverlust oder zunehmender sozialer Isolation am Arbeitsplatz, wohingegen Vertrauen eher mit der Nützlichkeit sowie dem Spaß im Umgang mit dem Roboter assoziiert wird. Somit könnte auch die Tatsache, dass die Framing-Texte eher die Ängste der Mitarbeitenden anstatt die Nützlichkeit des Cobots adressierten, zur stärkeren Veränderung des Misstrauensniveau

geführt haben. In diesem Fall wäre die beobachtete abweichende Entwicklung nicht als verallgemeinerbare Erkenntnis, sondern als spezifische Konsequenz der verwendeten Materialien in den Experimentalstudien zu sehen.

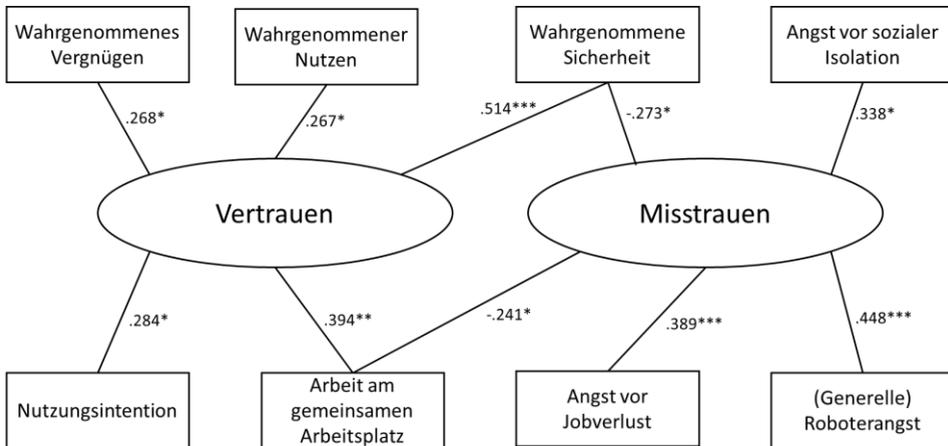


Abbildung 4.35: Grafische Darstellung der signifikanten Korrelationen einzelner Akzeptanzkonstrukte zu Vertrauen oder Misstrauen. Die Zahlenwerte repräsentieren die Korrelationskoeffizienten.

4.6.4.4 Vertrauensanstieg durch reale Interaktion

In den experimentellen Daten lässt sich deutlich erkennen, dass nach der realen MRI die Vertrauens- bzw. Misstrauenswerte jeweils signifikant höher bzw. tiefer lagen als davor. Zwischen diesen Messzeitpunkten zeigten Vertrauen und Misstrauen das erwartete anti-proportionale Verlaufsmuster. Dieser Befund steht in Einklang mit der Forschung, die praktischen Vorerfahrungen einen positiven Einfluss auf das Vertrauen und auf die Akzeptanz in Bezug auf Technologie zuschreibt (Sanders et al., 2017; Turja & Oksanen, 2019). Entsprechend beschrieben auch einige Versuchsteilnehmende in den Freitext-Kommentaren, dass sich anfänglich skeptische Gefühle schnell verflüchtigten und einem Gefühl der Vertrautheit wichen. Frühere Forschungen zeigten bereits, dass Personen einen Roboter mehr mochten, nachdem sie Kontakt zu diesem hatten (Wullenkord et al., 2016, S. 984). Dabei erwies es sich als irrelevant, ob diese Personen den Roboter anfassen konnten, ihn zu Gesicht bekamen oder nur über abstrakte Erfahrungen bspw. durch mediale Darstellungen verfügten (Müller-Abdelrazeq et al., 2019; Wullenkord et al., 2016). Wie bereits in Kapitel 4.5.4.3 erwähnt, wird dieser Zugewinn an Vertrauen auch in Bezug zu Robotern in der Regel mit dem aus der Sozialpsychologie bekannten *Effekt der Darbietungshäufigkeit* (*mere exposure effect*) (Bornstein, 1989) erklärt, der eine Familiarisierung mit dem Roboter zur Folge hat (Wullenkord et al., 2016). Grundsätzlich besagt der Effekt, dass anfangs neutral eingestellte Personen mit zunehmender Häufigkeit des Kontakts ihr Gegenüber

positiver wahrnehmen. Über die mentalen Prozesse, die diesen Effekt herbeiführen, existieren verschiedene Theorien (Montoya, Horton, Vevea, Citkowicz & Lauber, 2017). Einige Autor:innen haben diesen Effekt bereits auf den MRI-Kontext übertragen, um damit positivere Beurteilungen des Roboters über die Interaktionszeit hinweg zu begründen (de Graaf, Ben Allouch & van Dijk, 2016; Müller-Abdelrazeq, 2020, S. 42; Wullenkord et al., 2016). Die signifikanten Vertrauenssteigerungen durch den realen Kontakt mit einem Cobot in Experimentalstudie 2 erscheinen zunächst konform mit dem Effekt der Darbietungshäufigkeit. Allerdings stellt sich die Frage, ob die reine Darbietungshäufigkeit oder alternativ die Akkumulation dezidiert positiver Nutzungserfahrungen für die Vertrauenssteigerung ursächlich ist. So zeigte eine Studie zur Untersuchung der Persistenz von Effekten des unheimlichen Tals über die Dauer einer Interaktion, dass die Versuchsteilnehmenden einen als angenehm empfundenen Roboter nur nach einer positiv verlaufenen Interaktion mehr mochten (Złotowski, Sumioka et al., 2015). Aufschlussreich ist in diesem Zusammenhang die Analyse derjenigen empirischen Vertrauensverläufe aus Experimentalstudie 2, bei denen ein Fehler des Cobots das Eingreifen der Versuchsleitung erforderte. In diesen Fällen sollte sich demnach keine konsistente Akkumulation rein positiver Nutzungserfahrungen ergeben haben. Die Daten zeigen auch in diesen Fällen eine Vertrauenssteigerung, wenngleich in abgeschwächter Form. Zumindest zu einem gewissen Anteil scheint sich die Vertrauenssteigerung demnach durch den Effekt der Darbietungshäufigkeit begründen zu lassen.

Aufgrund der intensiven Mensch-Cobot-Interaktion inklusive kollaborierender Arbeitsschritte in Experimentalstudie 2, die über die reine Darbietung und Beobachtung eines Roboters hinausgeht, könnte der Effekt der Darbietungshäufigkeit besonders deutlich ausgefallen sein. Dabei ermöglichte die direkte Kollaboration mit dem Cobot auch haptische Reize in Form physischer Berührungen und erzeugte vermutlich ein verstärktes Gefühl der Vertrautheit. In Übereinstimmung damit zeigte sich in einer früheren Studie, dass die Versuchsteilnehmenden ein höheres Vertrauen in einen Roboter entwickelten, wenn sie diesen – wie in Experimentalstudie 2 – mittels eines Knopfdrucks selbst starteten (Ullman & Malle, 2017). Unklar ist hierbei, ob die Berührung selbst oder die Kontrollerfahrung über den Roboter dafür maßgeblich ist. Der Einfluss von Berührungen auf die Vertrauensbildung bedarf noch weiterer Untersuchungen im Kontext verschiedener Interaktionstypen (Law, Malle & Scheutz, 2021). Generell unterstreichen die Ergebnisse den in den Vorstudien geäußerten Wunsch von Produktionsmitarbeitenden nach realen Begegnungsmöglichkeiten und Experimentierräumen mit Cobots. Gleichwohl ist einschränkend zu erwähnen, dass der Effekt der Darbietungshäufigkeit gemäß dem aktuellen Forschungsstand nur unter bestimmten Bedingungen auftritt. Während der Effekt in den genannten MRI-Studien stets zur Erklärung von Vertrauenssteigerungen ins Feld geführt wird, geht die sozialpsychologische Forschung davon aus, dass der Effekt nur eintritt, wenn die beteiligten Personen zu Beginn der Interaktion ihrem Gegenüber bereits neutral oder positiv eingestellt sind (Kawakami, 2012). Insofern liegt die Vermutung nahe, dass die neutrale Grundhaltung der

Studierenden gegenüber Robotern (gemäß NARS-Skala) in Experimentalstudie 2 eine notwendige Bedingung dafür darstellte, dass die Versuchsteilnehmenden trotz der beobachteten Fehler des Roboters diesem nach der Interaktion mehr vertrauten.

4.6.5 Zwischenfazit

Entgegen der Erwartung erwies sich die experimentelle Manipulation im Gegensatz zu Experimentalstudie 1 als unwirksam. Zwar zeigte sich ein geringfügig höherer Wert bei der wahrgenommenen Menschenähnlichkeit für die Experimentalgruppe, die ein menschenähnliches Framing erhielt, allerdings waren die Unterschiede zur Versuchsgruppe mit maschinellem Framing nicht signifikant. Hypothetisch lässt sich dies auf ein Zusammenspiel aus geringerer Teststärke aufgrund der deutlich geringeren Stichprobenanzahl und einer geringeren Effektstärke des Framings bedingt durch die vermutlich robotererfahrenere Stichprobe der Studierenden zurückführen. Ferner könnte das experimentelle Setting in Form einer Präsenzstudie die Effektivität des Framings beeinflusst haben. Möglicherweise bewirkte die unmittelbar bevorstehende Interaktion, dass sich die Versuchspersonen weniger intensiv mit den Texten beschäftigten und affektive Vertrauenskomponenten stärker zur Geltung kamen. Eindrücklich beschrieben zwei Versuchspersonen, wie wenig die für sie offensichtlich vermenschlichenden Beschreibungen aus den Framing-Texten mit ihrem mentalen Modell von Robotern und dem Erleben der Interaktion in Einklang standen. Die damit verbundene kognitive Dissonanz könnte zu einer Abwehrhaltung gegenüber den vermenschlichenden Beschreibungen geführt haben.

Aufgrund der fehlgeschlagenen experimentellen Manipulation ließen sich die ursprünglichen Forschungsfragen nicht wie vorgesehen im Rahmen eines Vergleichs der ursprünglichen Versuchsgruppen beantworten. Zwar zeigte sich eine deskriptive Tendenz, dass die Versuchsteilnehmenden mit einem höheren initialen Vertrauen die reale Interaktion *ex post* positiver bewerteten. Allerdings ergaben sich keine signifikanten Unterschiede zu den Teilnehmenden mit unterdurchschnittlichem Vertrauen. Überdies basiert dieser Vergleich auf einer nachträglichen Gruppeneinteilung, da das initiale Vertrauensniveau nicht durch das Framing experimentell manipuliert werden konnte. Somit liegt nur eine schwache, auf einem quasi-experimentellen Versuchsdesign basierende Evidenz zur Beantwortung von *FF4* vor, gemäß derer ein höheres einstellungsbasiertes Vertrauen zu einer positiveren Wahrnehmung einer Interaktion führt. Ebenfalls aufgrund der Tatsache, dass das Framing keine signifikanten Effekte hervorrief, lassen sich in Hinblick auf *FF5* keine Rückschlüsse auf die Persistenz solcher Effekte über die Interaktion hinaus ziehen. In Kapitel 5.4.4 werden daher alternativ theoriebasierte Ansätze zur Beantwortung dieser Frage präsentiert.

Die durchgeführte Korrelationsanalyse ergab Hinweise darauf, dass die grundsätzliche Einstellung gegenüber Robotern einen recht stabilen Prädiktor für das Vertrauensniveau

zu verschiedenen Messzeitpunkten darstellt. Diese Erkenntnis ist insofern naheliegend, als bei aller experimentell gefundenen Wirksamkeit bestimmter Effekte auf das Vertrauen nicht davon auszugehen war, dass diese im Rahmen des experimentellen Settings und binnen einer solch kurzen Zeitspanne eine fundamentale Änderung der Grundeinstellung zu bewirken vermögen. Ferner scheinen Vertrauen und Misstrauen mit unterschiedlichen Aspekten der Akzeptanz und verschiedenen Eigenschaften des Cobots assoziiert zu sein. Während Vertrauen in Zusammenhang mit positiven Faktoren wie Vergnügen und Nutzen steht, ergaben sich Korrelationen zwischen Misstrauen und diversen Ängsten. Das könnte erklären, warum sich das Misstrauen durch das Framing gruppenübergreifend signifikant reduzierte, während sich das Vertrauen nicht signifikant veränderte. In Hinblick auf den Einfluss des realen Kontakts mit einem Roboter ergab sich wiederum insofern ein konsistentes Bild, als das Vertrauen dadurch signifikant zu- und das Misstrauen signifikant abnahm.

Die Diskrepanz der Ergebnisse zwischen Experimentalstudie 1 und Experimentalstudie 2 wirft ein besonderes Schlaglicht auf die möglicherweise unterschiedliche Effektivität von Framing in Bezug auf verschiedene Personengruppen und Versuchssettings. Eine ange-dachte Nacherhebung, um die unterschiedliche Stichprobenanzahl als rein methodische Ursache für die stark abweichenden Ergebnisse von Experimentalstudie 1 und Experimentalstudie 2 auszuschließen, konnte aufgrund forschungspragmatischer Einschränkungen im Zusammenhang mit der Covid-19-Pandemie nicht durchgeführt werden. Eine solche weiterführende Untersuchung mit vergleichbarer Stichprobengröße erscheint allerdings in der Zukunft als sinnvoll. Um Rückschlüsse auf die Relevanz personenspezifischer Merkmale in Hinblick auf die unterschiedlichen Ergebnisse zu erhalten, werden die unterschiedlichen Populationen der beiden Experimentalstudien im folgenden Kapitel einem Vergleich unterzogen.

4.7 Gegenüberstellung der Experimentalstudien

4.7.1 Vergleich der Stichproben

Die Ergebnisse des in Abbildung 4.36 dargestellten Vergleichs zwischen den Produktions-mitarbeitenden aus Experimentalstudie 1 und den Studierenden aus Experimentalstudie 2 offenbaren signifikante Unterschiede in sämtlichen erhobenen populationsspezifischen Parametern. Mit Ausnahme der eingeschätzten Menschenähnlichkeit des Cobots wurden sämtliche Merkmale jeweils zu Beginn der Experimentalstudie gemessen, sodass sich eine Beeinflussung durch eine experimentelle Intervention ausschließen lässt. Die teilnehmenden Studierenden zeigten sich Robotern gegenüber weniger negativ eingestellt, vertrauten dem vorgestellten Cobot initial mehr, verfügten über etwas mehr Vorkenntnisse und waren

stärker von ihrer Technikkompetenz überzeugt. Ferner nahmen sie den Cobot nach dem Framing insgesamt signifikant weniger menschenähnlich wahr, was auf ein ausgeprägtes technikorientiertes mentales Modell von Robotern hindeutet.

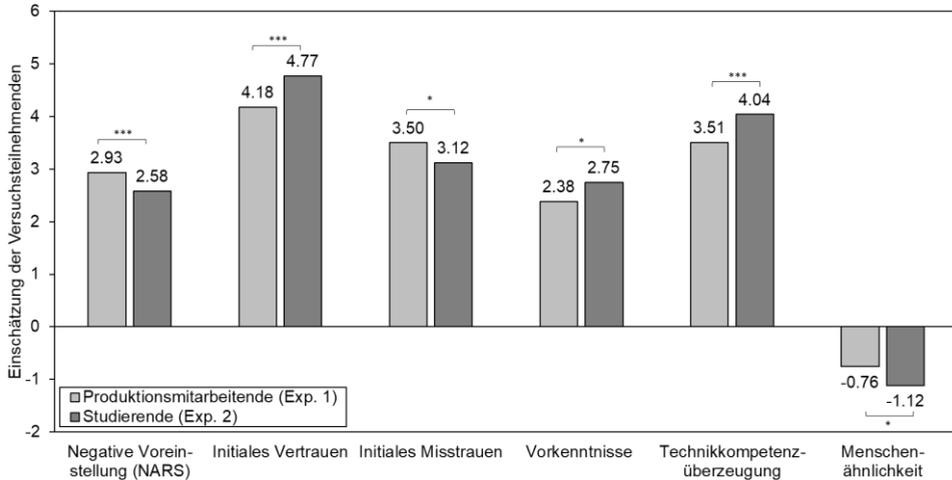


Abbildung 4.36: Vergleich verschiedener Einstellungen und Einschätzungen in Bezug auf Cobots von Produktionsmitarbeitenden aus Experimentalstudie 1 und Studierenden aus Experimentalstudie 2. Bei der Darstellung ist zu beachten, dass die einzelnen Variablen auf unterschiedlichen Skalen gemessen wurden.

Die Geschlechtsverteilung fiel in beiden Studien sehr ähnlich aus, wohingegen sich das Durchschnittsalter erwartungsgemäß signifikant unterschied, wie die detaillierten inferenzstatistischen Ergebnisse in Tabelle 35 verdeutlichen. Überdies differierten die praktischen arbeitsbezogenen Vorerfahrungen mit Robotern, die 27.9 % der Studierenden aufwiesen, bei den Produktionsmitarbeitenden allerdings durch die Screening-Kriterien gezielt ausgeschlossen waren.

Tabelle 35: Inferenzstatistischer Vergleich populationsbezogener Kennwerte aus Experimentalstudie 1 mit Produktionsmitarbeitenden und aus Experimentalstudie 2 mit Studierenden.

	<i>t</i> -Test ¹³		
	<i>df</i>	<i>t</i>	<i>p</i>
Negative Voreinstellung zu Robotern (NARS)	193.3	4.902	< .001***
Initiales Vertrauen (vor dem Framing)	215	-3.575	< .001***

¹³ Aufgrund einer Stichprobenanzahl pro Gruppe von $n > 50$ wurden aus Gründen der besseren Vergleichbarkeit einheitlich *t*-Tests zur Überprüfung auf signifikante Unterschiede eingesetzt, auch wenn keine Normalverteilung der Daten vorlag.

Initiales Misstrauen (vor dem Framing)	215	2.753	.012*
Vorkenntnisse zu Robotern	215	-2.438	.016*
Technikkompetenzüberzeugung	152.3	-4.227	< .001***
Wahrgenommene Menschenähnlichkeit	157.8	2.262	.025*
Alter	183.6	22.206	< .001***

Demnach liegt die Vermutung nahe, dass sich die unterschiedlichen Befunde zur Effektivität von Framing aus den beiden Experimentalstudien zumindest teilweise auf die stichprobenspezifischen Unterschiede zurückführen lassen. Eine weniger intensive Beschäftigung mit Robotern, eine geringere Technikkompetenzüberzeugung und das Fehlen praktischer Erfahrungen mit Robotern bei den Produktionsmitarbeitenden aus Experimentalstudie 1 deuten auf weniger gefestigte mentale Modelle und damit eine stärkere Beeinflussbarkeit durch Framing hin. Einschränkend ist allerdings zu erwähnen, dass die Datenlage keine Rückschlüsse darauf zulässt, welche Persönlichkeitsmerkmale zu welchem Grad die Wirksamkeit determinieren, da viele Faktoren stark konfundieren. Dies gilt bspw. für die Eigenschaft, sich im Studium, anstatt im Arbeitsleben zu befinden, die stark mit weiteren Merkmalen wie dem Alter und dem bisherigen Bildungsabschluss zusammenhängt. Außerdem kann ein Einfluss des experimentellen Settings nicht ausgeschlossen werden. Gleichwohl verdeutlichen die signifikanten Unterschiede in der Gesamtschau, dass Studierende keine repräsentative Population für Produktionsmitarbeitende darstellen, also für diejenige Personengruppe, die perspektivisch in der Praxis mit Cobots interagieren wird. Vor diesem Hintergrund ist die externe Validität und damit die Aussagekraft der Ergebnisse von Studien als problematisch einzustufen, die auf Studierende als Stichprobe zurückgreifen und gleichzeitig auf Produktionsmitarbeitende generalisieren (vgl. hierzu Elprama et al., 2016, S. 1; Horstmann et al., 2018, S. 20; Müller-Abdelrazeq et al., 2019, S. 115; Richert et al., 2017, S. 79; Richert, 2018, S. 109).

4.7.2 Vergleich der Vertrauensentwicklungen

Abbildung 4.37 zeigt anhand der experimentellen Daten aus beiden Experimentalstudien die unterschiedlichen Vertrauensniveaus der Versuchsgruppen zu den jeweiligen Messzeitpunkten. In Einklang mit den Eindrücken aus den qualitativen Studien unterscheidet sich das initiale Vertrauen zwischen den verschiedenen Versuchsteilnehmenden mitunter stark. Offenbar verfügten diese über sehr unterschiedliche mentale Modelle von der bevorstehenden MRI, die sich womöglich aus der Übertragung früherer Erfahrungen mit ähnlichen Geräten speisten (Kaplan, Kessler et al., 2021, S. 145). Ferner ist davon auszugehen, dass die Varianz im initialen Vertrauen auch auf Unterschiede in der individuellen Vertrauensdisposition zurückzuführen ist, insbesondere da zu diesem Zeitpunkt noch keine Informationen zu Cobots vermittelt worden waren (Alarcon et al., 2021, S. 161). In einem

freiwilligen Nutzungskontext wie dem Konsumgüterbereich wäre das initiale Vertrauen in erster Linie für die Nutzungsentscheidung bedeutsam. Am Arbeitsplatz als verpflichtender Nutzungskontext wirkt sich das initiale Vertrauensniveau auf die Angemessenheit der Erwartungen und damit auf die Wahrnehmung sowie die Effizienz der Interaktion aus (H. Choi & Swanson, 2021). Ein hohes Vertrauensniveau *a priori* geht dabei tendenziell mit einer positiveren Wahrnehmung einer realen MRI einher, wie die Ergebnisse aus Experimentalstudie 2 zeigen (vgl. Kapitel 4.6.4.2).

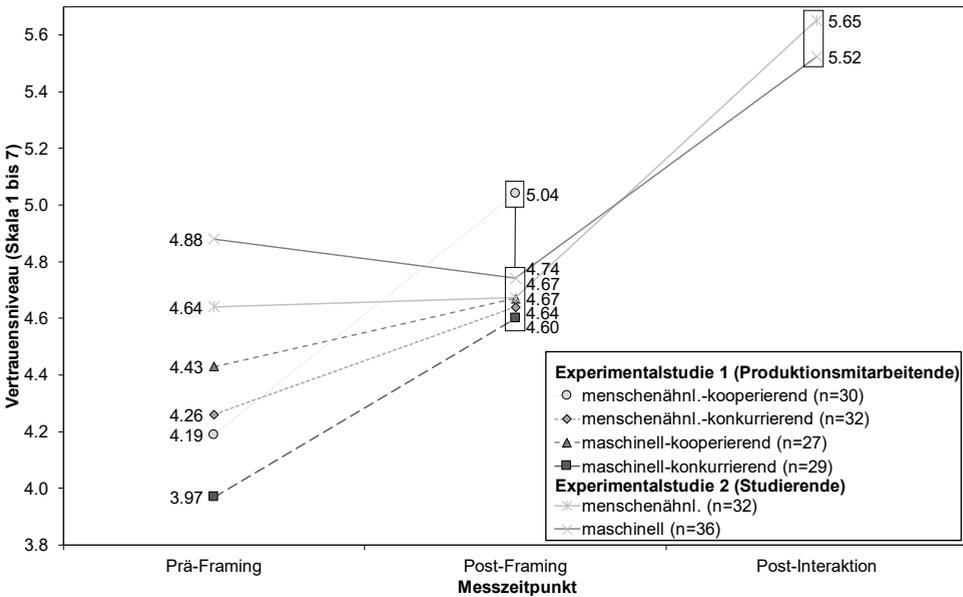


Abbildung 4.37: Vertrauensentwicklung in allen Experimentalgruppen beider Experimentalstudien über die Messzeitpunkte hinweg. Die Daten aus der zweiten Experimentalstudie sind mit durchgezogenen Linien dargestellt.

Ferner verdeutlicht Abbildung 4.37, dass die Vertrauenswerte nach dem Framing auf ein vergleichsweise einheitliches Niveau konvergierten. Das lässt sich darauf zurückführen, dass das Vertrauensniveau bei den Produktionsmitarbeitenden, die an Experimentalstudie 1 teilnahmen, aufgrund des Effekts des imaginierten Kontakts anstieg (vgl. Kapitel 4.5.4.3), wohingegen das Niveau bei den Studierenden aus Experimentalstudie 2 weitestgehend unverändert blieb (vgl. vertiefende Diskussion in Kapitel 4.7.3). Die variierenden Vertrauenssteigerungen in den verschiedenen Versuchsgruppen aus Experimentalstudie 1 lassen sich auf den originären Effekt des Framings zurückführen, d. h. auf die Unterschiede in den Formulierungen innerhalb der jeweiligen Framing-Texte. Besonders augenfällig ist das erhöhte Vertrauensniveau und die überdurchschnittlich starke Vertrauenssteigerung in der menschenähnlich-kooperierenden Versuchsbedingung (vgl.

Kapitel 4.5.4.2). Außerdem verdeutlicht das Diagramm den Effekt des realen Kontakts aus der zweiten Experimentalstudie, der zu einem deutlichen Anstieg der Vertrauenswerte zwischen den Messzeitpunkten *Post-Framing* und *Post-Interaktion* führte.

4.7.3 Einfluss des imaginierten und realen Kontakts

Wie bereits in den spezifischen Diskussionen der Experimentalstudien beschrieben, liegt aus früheren Studien eine gewisse Evidenz vor, dass neben einem realen Kontakt offenbar auch ein intensiv vorgestellter Kontakt mit einem Roboter im Sinne einer mentalen Simulation, also einer dynamischen Anwendung eines mentalen Modells, zu einer Vertrauenssteigerung führen kann (Wullenkord & Eyssel, 2014; Wullenkord et al., 2016; Wullenkord & Eyssel, 2019). Da die Konstruktion mentaler Modelle als Methode des Textverstehens gilt (vgl. Kapitel 3.1.1), liegt die Vermutung nahe, dass solche mentalen Vorstellungen bei der Rezeption von Texten automatisch entstehen, wenngleich diese nicht notwendigerweise visuell vor dem inneren Auge sichtbar werden. Diese bis dato noch wenig empirisch abgesicherte Hypothese lässt sich auf Basis der Ergebnisse aus Experimentalstudie 1 mit den Produktionsmitarbeitenden stützen. Ein komplexeres Bild ergibt sich bei ergänzender Betrachtung der Ergebnisse aus Experimentalstudie 2 unter Rückgriff auf die Differenz zwischen Vertrauen und Misstrauen. In dieser Studie verringerte sich das Vertrauen durch den imaginierten Kontakt nicht signifikant, wohingegen das Misstrauen signifikant abnahm.

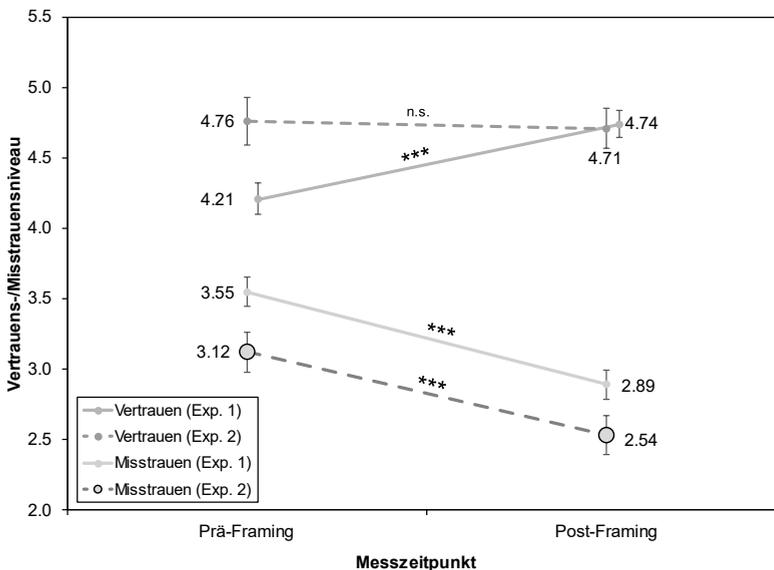


Abbildung 4.38: Vertrauens- und Misstrauensentwicklung zwischen den Messzeitpunkten in den jeweiligen Experimentalgruppen in Experimentalstudie 1 und Experimentalstudie 2.

Die in Abbildung 4.38 visualisierten Verlaufsmuster aus den beiden Experimentalstudien erlauben zwei unterschiedliche Erklärungshypothesen. Einerseits wäre es möglich, dass der Effekt des imaginierten Kontakts nur auf Misstrauen wirkt. Die positive Vertrauensentwicklung in Experimentalstudie 1 müsste dann durch einen andersartigen Effekt erklärt werden. Dieser Erklärungsansatz erscheint jedoch wenig plausibel und lässt sich nicht durch entsprechende Literatur stützen. Andererseits könnte der Effekt des imaginierten Kontakts sowohl das Vertrauen als auch das Misstrauen beeinflussen. In diesem Fall wäre allerdings zu begründen, weshalb dieser Effekt in Experimentalstudie 2 nur in Bezug auf das Vertrauen ausbleibt. In diesem Kontext ist sowohl eine methodische als auch eine inhaltliche Begründung denkbar. Aus methodischer Sicht ist zu bedenken, dass die initialen Vertrauenswerte in Experimentalstudie 2 mit $M = 4.76$ deutlich höher als in Experimentalstudie 1 ($\Delta_M = 0.55$) lagen, wohingegen sich die Werte für das initiale Misstrauen auf deutlich niedrigerem Niveau im mittleren Bereich der siebenstufigen Skala bewegten ($M = 3.12$) und sich weniger stark zu Experimentalstudie 1 unterschieden ($\Delta_M = 0.43$). Möglicherweise ließen sich Steigerungen eines ohnehin schon hohen Vertrauensniveaus angesichts einer ausgeprägten Tendenz zur Mitte im Antwortverhalten nur noch unzureichend auf der Skala abbilden. Eine relevante Kontextbedingung für den geringen Effekt des imaginierten Kontakts in Experimentalstudie 2 stellt in diesem Zusammenhang die Stichprobe der teilnehmenden Studierenden dar. Diesen waren MRI-Situationen möglicherweise ohnehin aus ihrem Studienalltag oder ihrem privaten Interessensbereich geläufig. Entsprechend verfügten sie bereits über mentale Modelle, um den Cobot zu repräsentieren, und über vorgefertigte Handlungsskripte für die bevorstehende Interaktion, sodass die Lektüre der Texte und die damit potenziell verbundene mentale Simulation keine Veränderungen in den bestehenden kognitiven Strukturen bewirkte. Das könnte auch erklären, weshalb das initiale Vertrauensniveau der Studierenden *vor* dem Framing in Experimentalstudie 2 auf vergleichbarem Niveau wie das Vertrauen der Produktionsmitarbeitenden *nach* dem Framing in Experimentalstudie 1 lag. Gleichwohl stellt diese Beobachtung nur ein vergleichsweise schwaches Indiz dar.

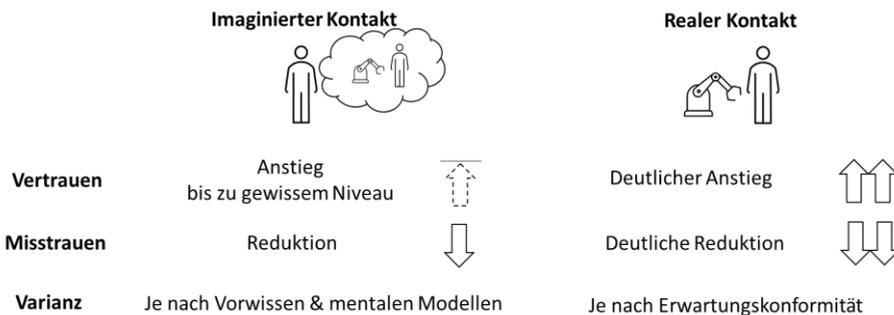


Abbildung 4.39: Schematische Auswirkungen von imaginiertem und realem Kontakt auf Vertrauen und Misstrauen.

Abgesehen von diesem Erklärungsansatz ließe sich aus inhaltlicher Perspektive auf Basis der Daten schlussfolgern, dass sich Vertrauen im Vorfeld einer Interaktion nur bis zu einem gewissen Grad steigern lässt, wohingegen Misstrauen in höherem Maße abgebaut werden kann. Möglicherweise spielen in diesem Zusammenhang die in Kapitel 4.6.4.3 erläuterten unterschiedlichen Assoziationen von Vertrauen und Misstrauen eine Rolle. Hypothetisch könnte der imaginierte Kontakt zwar zu einem Abbau bestehender Ängste in Bezug auf die Roboternutzung und zu einer Verringerung des Misstrauensniveaus führen, aber nicht imstande sein, Vertrauen über ein gewisses Niveau hinaus aufzubauen, solange tatsächliche Nutzungserfahrungen in der realen Umgebung fehlen. Das weckt Assoziationen zur *Zwei-Faktoren-Theorie der Arbeitsmotivation* nach Herzberg, Mausner und Snyderman (1959/2011), die zwischen Hygiene- und Motivationsfaktoren unterscheidet. Der Misstrauensabbau käme demnach einem bloßen Hygienefaktor gleich, der eine prinzipielle Technologienutzung ermöglicht, aber im Gegensatz zum Vertrauensaufbau noch keine engagierte, neugierige und kooperationsorientierte Interaktion begünstigt. Abbildung 4.39 visualisiert die vermuteten Effekte von imaginiertem und realem Roboterkontakt in Bezug auf Mensch-Cobot-Vertrauen und -Misstrauen.

Außerdem lässt sich auch an dieser Stelle nicht ausschließen, dass der Unterschied zwischen Online- und Präsenzsetting in den beiden Experimentalstudien für die divergierenden Ergebnisse verantwortlich ist. So könnten die Studierenden aufgrund der bekanntermaßen bevorstehenden realen Interaktion den beschreibenden Framing-Texten eine stark untergeordnete Relevanz beigemessen und damit deren Wirkung unterminiert haben. Das lässt sich allerdings nur vermuten, da in den Experimentalstudien nicht explizit abgefragt wurde, wie intensiv sich die Versuchsteilnehmenden mit dem Text auseinandergesetzt und sich eine MRI-Situation vorgestellt hatten. Allerdings liefert auch diese Hypothese keine hinreichende Erklärung dafür, warum sich das Misstrauensniveau in beiden Experimenten über alle Bedingungen hinweg durch das Framing signifikant veränderte. Für eine genauere Erklärung der Effekte und eine Identifikation der Wirkmechanismen sind weitere Untersuchungen notwendig, die auch die subjektive Bewertung der Interaktion einbeziehen (vgl. dazu die Ergebnisse von Zlotowski, Sumioka et al., 2015). Das gilt insbesondere, da die experimentellen Daten keine Rückschlüsse darauf zulassen, inwiefern eine womöglich sogar unbewusste Imagination der bevorstehenden Interaktion in den jeweiligen Experimentalstudien und Versuchsbedingungen stattgefunden hat. Zu beachten ist hierbei ferner, dass die Versuchspersonen in den durchgeführten Experimentalstudien zwar mit einem beschreibenden Text konfrontiert, aber im Gegensatz zu bestehenden Studien zum imaginierten Kontakt nicht aktiv zur Imagination aufgefordert wurden.

4.8 Zusammenfassung der Ergebnisse

Das Erkenntnisinteresse der vorliegenden Dissertation wurde durch die in Kapitel 3.5 formulierten Forschungsfragen umrissen. Tabelle 36 fasst die zentralen Erkenntnisse aus den empirischen Untersuchungen in Bezug auf diese Forschungsfragen zusammen und stellt zusätzlich explorative Erkenntnisse (*incidental findings*) dar, die sich im Verlauf der Forschung zusätzlich ergeben haben.

Tabelle 36: Zentrale Erkenntnisse aus den verschiedenen empirischen Untersuchungen in Hinblick auf die Forschungsfragen sowie davon unabhängige explorative Erkenntnisse.

<p><i>FF1</i>: Welche Rolle spielt Vertrauen als Erfolgsfaktor für betriebliche Cobot-Einführungen aus Sicht von Unternehmensvertreter:innen im Bereich der industriellen MRI?</p> <ul style="list-style-type: none"> • <i>Vorstudien</i>: Produktionsmitarbeitende auf leitender und operativer Ebene beurteilen initiales und dynamisches Vertrauen als relevant für den Erfolg einer Cobot-Einführung. • <i>Vorstudien</i>: Produktionsmitarbeitende auf leitender und operativer Ebene erachten die Relevanz von anthropomorphem Design und Vermenschlichung bei expliziter Befragung als gering, zeigen allerdings (wie auch in Exp. 1) impliziten Anthropomorphismus. • <i>Vorstudien</i>: Produktionsmitarbeitende auf operativer Ebene übertragen verschiedene mentale Modelle von Menschen oder von klassischer Automatisierung auf Cobots und reagieren sensibel auf Umschreibungen der Mensch-Cobot-Relation. • <i>Vorstudien</i>: Ängste vor Arbeitsplatzverlust sind unter Produktionsmitarbeitenden auf operativer Ebene virulent und werden (auch auf der Management-Ebene) als sehr kritisch für den Erfolg einer Cobot-Einführung empfunden.
<p><i>FF2</i>: Inwiefern lässt sich die Wahrnehmung eines Cobots und der Mensch-Cobot-Relation mittels sprachlichen Framings beeinflussen?</p> <ul style="list-style-type: none"> • <i>Exp. 1 (Produktionsmitarbeitende)</i>: Bereits kurze Beschreibungen von Cobots mit wenigen sprachlichen Variationen (Framing) beeinflussen signifikant die wahrgenommene Menschenähnlichkeit und die wahrgenommene Kooperationsorientierung der Mensch-Cobot-Relation. • <i>Exp. 2 (Studierende)</i>: Dieselben Framing-Texte zur Manipulation der wahrgenommenen Menschenähnlichkeit wie in Experimentalstudie 1 erzielen keine Wirkung. • <i>Vergleich Exp. 1 & Exp. 2</i>: Die Wirkung des Framings hängt vermutlich stark vom Rezipienten ab (siehe <i>FF6</i>).

<p><i>FF3</i>: Wie wirken sich die wahrgenommene Menschenähnlichkeit eines Cobots und die wahrgenommene Kooperationsorientierung der Mensch-Cobot-Relation auf das Mensch-Cobot-Vertrauen aus?</p> <ul style="list-style-type: none"> • <i>Exp. 1</i>: Die Wirkung von Vermenschlichung auf Vertrauen wird durch die Mensch-Cobot-Relation moderiert. Wenn die Relation als kooperationsorientiert empfunden wird, wirkt sich ein vermenschlichendes Framing signifikant positiver auf das Vertrauen aus als eine maschinelle Umschreibung des Cobots.
<p><i>FF4</i>: Inwiefern wirken sich durch das Framing induzierte Veränderungen am initialen Vertrauen auf die reale Interaktion aus?</p> <ul style="list-style-type: none"> • <i>Exp. 2</i>: Es besteht ein Zusammenhang zwischen einem hohen Vertrauen in den Cobot und einer positiven Wahrnehmung einer Interaktion mit demselben (schwache Evidenz).
<p><i>FF5</i>: Inwiefern überdauern Framing-Effekte eine reale MRI? – keine Ergebnisse aufgrund unwirksamer experimenteller Manipulation in Experimentalstudie 2 –</p>
<p><i>FF6</i>: Inwiefern reagieren verschiedene Populationen unterschiedlich auf das sprachliche Framing?</p> <ul style="list-style-type: none"> • <i>Vergleich Exp. 1 & Exp. 2</i>: Studierende (vorwiegend aus technischen Studiengängen) unterscheiden sich in zahlreichen Merkmalen signifikant von Produktionsmitarbeitenden (ohne praktische Vorerfahrung mit Robotern). Sie verfügen über eine (jeweils signifikant) weniger negative Grundeinstellung gegenüber Robotern, ein höheres initiales Vertrauen, ein geringeres initiales Misstrauen, mehr Vorkenntnisse in Bezug auf Roboter und eine höhere Technikkompetenzüberzeugung. • <i>Vergleich Exp. 1 & Exp. 2</i>: Framing wirkt sich im Gegensatz zu den untersuchten Produktionsmitarbeitenden (ohne praktische Vorerfahrung mit Robotern) nicht auf die untersuchten Studierenden (vorwiegend aus technischen Studiengängen) aus.
<p>Explorative Erkenntnisse:</p> <ul style="list-style-type: none"> • <i>Exp. 1 & Exp. 2</i>: Die Einstellung gegenüber Robotern ist ein recht stabiler Vertrauensindikator. • <i>Exp. 1 & Exp. 2</i>: Vertrauen und Misstrauen zeigen zumeist antiproportionale Verlaufsmuster. Misstrauen kann durch unspezifische Informationen über Roboter vermutlich stärker beeinflusst werden als Vertrauen. Unterschiede in der wahrgenommenen Menschenähnlichkeit und der Mensch-Cobot-Relation wirken sich vergleichsweise weniger stark auf Misstrauen als auf Vertrauen aus. • <i>Exp. 1 (Produktionsmitarbeitende)</i>: In allen Experimentalgruppen kommt es zu einer Vertrauenssteigerung (in drei von vier Gruppen signifikant) durch die Lektüre der Cobot-Beschreibungen (Effekt des imaginierten Kontakts).

- *Exp. 2 (Studierende)*: Durch die tatsächliche Interaktion mit dem Cobot kommt es zu starken Vertrauenssteigerungen (Effekt der Darbietungshäufigkeit).
- *Exp. 1 & Exp. 2*: Die Konzepte der Akzeptanz und des Vertrauens zeigen starke Korrelationen. Verhaltensbezogene Konsequenzen von einstellungsbasiertem Vertrauen sind naheliegend.

5 Einbettung der Ergebnisse in die aktuelle Forschung

Im Folgenden werden die Ergebnisse anhand zentraler Themenblöcke vertiefend diskutiert und in die aktuelle Forschung eingebettet, um der weiterhin wachsenden Dynamik des Forschungsfeldes Rechnung zu tragen. Die Themenblöcke orientieren sich im Wesentlichen an der Untergliederung des theoretischen Kapitels zur kognitiven MRI (Kapitel 3) und enthalten punktuelle Vertiefungen bei denjenigen Sachverhalten, die sich empirisch als besonders relevant erwiesen haben. Die skizzierten Erklärungsansätze unternehmen den Versuch, Theorieansätze und Denkfiguren aus verschiedenen Forschungsdisziplinen auf den interdisziplinären MRI-Kontext zu übertragen und damit produktive Anknüpfungspunkte für künftige Forschung zu identifizieren. Deren Ziel sollte darin bestehen, die skizzierten Ansätze stärker zu integrieren, zu umfassenden Theorien auszuarbeiten und einer hypothesengeleiteten empirischen Prüfung zu unterziehen.

5.1 Mensch-Cobot-Vertrauen am industriellen Arbeitsplatz

5.1.1 Phasen der dynamischen Vertrauensentwicklung

Wie insbesondere aus der ersten durchgeführten Vorstudie hervorgeht, schätzen Produktionsmitarbeitende auf Leitungsebene die Relevanz von Vertrauen für den Erfolg einer Cobot-Einführung sowohl vor als auch während der Interaktion als hoch ein (vgl. Kapitel 4.2.4). Die Differenzierung zwischen initialem und dynamischem Vertrauen sowie die Verwendung von Rückkopplungsschleifen in den gängigen Vertrauensmodellen (vgl. Kapitel 3.4.7 zu verschiedenen Vertrauensmodellen) deutet bereits die zeitliche Dynamik der Vertrauensentwicklung an. Diese wird maßgeblich durch die sog. Erwartungskongruenz beeinflusst. Diese beschreibt den Grad der Übereinstimmung der Erwartungen an die vertrauensnehmende Instanz, die sich aus dem initialen Vertrauen ergeben, mit den Beobachtungen während einer MRI. Die aktuelle Forschung nimmt zunehmend eine zeitdynamische Perspektive auf Vertrauen ein, indem sie das Ziel einer nachhaltig positiven Vertrauensentwicklung aufgrund von hoher Erwartungskongruenz im Gegensatz zu einer rein akzeptanzbeschaffenden initialen Vertrauenssteigerung akzentuiert (Nam & Lyons, 2021). Eine nachhaltig positive Vertrauensentwicklung basiert dabei auf einem angemessenen initialen Vertrauensniveau, aus dem sich realistische Erwartungen ableiten lassen.

Die Frage nach einem angemessenen initialen Vertrauen ist wiederum stets in Relation zur tatsächlichen Vertrauenswürdigkeit der vertrauensnehmenden Instanz zu beantworten (O'Neill, 2018), die die Erfahrungen und Beobachtungen während der Interaktion vorbestimmt. Eine exakte Festlegung der Vertrauenswürdigkeit setzt die Verfügbarkeit realistischer Daten über die tatsächlichen Fähigkeiten und Einschränkungen eines Roboters voraus (Schaefer, Perelman, Gremillion, Marathe & Metcalfe, 2021, S. 272).

Die Entwicklung des initialen Vertrauens beginnt bereits im Vorfeld der Interaktion. In dieser Phase ist neben dem dispositionalen und situativen Vertrauen das initial vermittelte Vertrauen relevant, das gemäß den Ergebnissen aus Experimentalstudie 1 u. a. von der Art der Informationsvermittlung, also dem sprachlichen Framing, abhängt. Gleichwohl existieren vielfältige weitere Komponenten und Einflussfaktoren, die mitunter ebenfalls auf subtile Art und Weise wirken und sich der Introspektion entziehen, sodass ein erschöpfender Überblick unmöglich ist (Nam et al., 2021, S. 479). Mit dem Übergang in die Phase der tatsächlich stattfindenden MRI steigt die Relevanz von Verhaltensbeobachtungen und Interaktionserfahrungen als Einflussfaktoren auf das dynamische Vertrauen (Abubshait & Wiese, 2017). Im Verlauf der Interaktion wirken sich mensch-, roboter- und kontextbezogene Einflussfaktoren mit je unterschiedlichem Gewicht auf die Vertrauensbildung aus (Kaplan, Kessler et al., 2021). Gleichzeitig gewinnen zusätzliche Persönlichkeitsfaktoren an Relevanz, wie bspw. die individuelle Bereitschaft, das gegenwärtige mentale Modell an Erfahrungen anzupassen (H. Choi & Swanson, 2021). Trotz der Vielfalt an zeitlich wirkenden Einflussfaktoren zeigte sich in den Experimentalstudien allerdings auch, dass die initiale Einstellung gegenüber Robotern einen relativ stabilen Prädiktor für das empfundene Vertrauen an allen Messzeitpunkten darstellt (vgl. Kapitel 4.6.3.6).

Die Entwicklung des dynamischen Vertrauens ist zu Beginn einer Interaktion sehr vulnerabel und kann starke Einbrüche erleiden (vgl. Kapitel 3.4.1). Frühe Negativerfahrungen können zu einem hohen Misstrauen und einer (passiven) Inakzeptanz führen und damit weitere unvoreingenommene (Positiv-)Erfahrungen verhindern (vgl. Kapitel 3.4.10). Auch sozialpsychologische Erkenntnisse im Zusammenhang mit dem Negativitätsbias (*negativity bias*) und dem Übereinstimmungsbias (*conformability bias*) lassen sich mit dieser Beobachtung in Verbindung bringen. Die Kernaussage besteht darin, dass als negativ bewertete Erfahrungen stärker wirken als positive Erfahrungen und bereits wenige Negativerfahrungen dazu führen, dass Personen ihre positiven Eigenschaftszuschreibungen überdenken und umkehren. Gleichsam ist eine Vielzahl positiver Erlebnisse notwendig, um negative Eigenschaftszuschreibungen zu korrigieren (Malle et al., 2020, S. 14; Rothbart & Park, 1986; Tausch, Kenworthy & Hewstone, 2007). Die qualitativen Aussagen der Produktionsmitarbeitenden aus Vorstudie 2 verweisen darauf, dass in dieser vulnerablen Phase insbesondere die (wahrgenommene) Zuverlässigkeit des Cobots unter besonderer Beobachtung steht. Schon begrifflich wird deutlich, dass die Zuverlässigkeit ein Maß dafür darstellt, inwiefern sich eine Person auf eine Technologie verlassen und ihr

vertrauen kann. Vertrauen kann dadurch enttäuscht werden, dass die positive Zukunftserwartung einer verlässlichen Funktionsweise nicht eintritt. Der Beobachtung von unerwarteten Ereignissen wie etwa Fehlern eines Cobots kommt demnach bei der Entwicklung des dynamischen Vertrauens eine hohe Bedeutung zu. Die Wahrnehmung einer hohen Inkongruenz zwischen Erwartungen und Erfahrungen stellt bestehende mentale Modelle in Frage und erzeugt dadurch kognitive Komplexität. Als symbolische Vertrauensverletzungen mit besonders starken Auswirkungen (Luhmann, 1968/2014, S. 735) lassen sich solche Ereignisse auffassen, die in scharfem Kontrast zum gegenwärtigen mentalen Modell der MRI-Situation stehen. Sie beschädigen zugleich das Zutrauen in eigene Erwartungen (vgl. Kapitel 3.4.1), stellen also in Frage, ob die jeweilige Person in der Lage ist, möglichst realitätsgetreue mentale Modelle zu entwickeln sowie präzise und sich bewahrheitende Zukunftserwartungen daraus abzuleiten. Wenn sich im Verlauf der MRI hingegen konsistente und erwartungskonforme Erfahrungen akkumulieren lassen, verfeinert sich das mentale Modell und die Anzahl verfügbarer Anhaltspunkte zur Ableitung von Zukunftserwartungen steigt. Dieses verstärkte Zutrauen in die eigene Vorhersagekraft führt zum Übergang in eine Phase der allmählichen Stabilisierung des Vertrauensniveaus. Dieser Übergang vollzieht sich fließend und lässt sich zeitlich nicht exakt bestimmen, da sich die Stabilisierung nicht in Abhängigkeit von der vergangenen Nutzungszeit einstellt, sondern sich an der subjektiven Interpretation der bisherigen Erfahrungen orientiert.

Im Rahmen der vorliegenden Dissertation ließen sich insbesondere empirische Belege für die Effekte des Framings sowie des imaginierten und realen Roboterkontakts identifizieren. Auf Basis aktueller Forschung lassen sich ferner Effekte der Erwartungsinkongruenz und des Misstrauens bzw. der passiven Inakzeptanz ableiten. Wenngleich sich die experimentellen Ergebnisse basierend auf einer kurzen einmaligen Interaktion nicht ohne Weiteres auf eine dauerhafte und alltägliche MRI über Wochen, Monate und Jahre hinweg verallgemeinern lassen, wird in Abbildung 5.1 der Versuch unternommen, mögliche zeitliche Vertrauensverläufe und auf diese unterschiedlich einwirkende Effekte zu skizzieren, um deren mögliches Zusammenspiel zu verdeutlichen. Einschränkend sei erwähnt, dass es sich um fiktive und schematische, explizit nicht empirisch erhobene Verlaufsskizzen handelt.

Verlaufsskizze A kennzeichnet den Nachteil eines niedrigen initialen Vertrauensniveaus, das zu einer passiven Inakzeptanz und damit zur Verhinderung positiver Nutzungserfahrungen führen kann. Kurvenverlauf B verdeutlicht, wie sich das initiale Vertrauen durch frühzeitige und geeignete Framing-Interventionen auf ein angemessenes Niveau erhöhen lässt, das zu ersten gelingenden Interaktionen führt und durch den vertrauenssteigernden Effekt des realen Erstkontaktes weiter ansteigt.

In welchem Ausmaß erwartungsinkongruente Beobachtungen (z. B. unerwartete Fehler des Cobots) ein unangemessen hohes Vertrauen zu beschädigen und in ein zu geringes Vertrauen zu transformieren vermögen, zeigt Kurvenverlauf C. Nach dem Einbruch des Vertrauens stabilisiert sich dieses zwar und kann durch die Akkumulation positiver Nutzungserfahrungen wieder langsam ansteigen, allerdings ohne ein angemessenes Niveau zu erreichen. Ein niedriges initiales Vertrauen oder frühe Enttäuschungen des dynamischen Vertrauens können damit die Vertrauensbildung langfristig begrenzen (Malle et al., 2020, S. 14). Kurvenverlauf D verdeutlicht hingegen, dass Framing-Interventionen nicht einseitig als Werkzeug zur kurzfristigen Vertrauenssteigerung zu sehen sind, sondern auch dem Abbau von überhöhtem initialem Vertrauen dienen und damit zu einem angemessenen und langfristig vergleichsweise stabilen Vertrauensniveau beitragen können. Zusammenfassend wird deutlich, dass der Bestimmung eines angemessenen Vertrauensniveaus eine große Wichtigkeit zukommt und sich die Eignung einzelner Maßnahmen nur vor dem Hintergrund der jeweils vorliegenden Diskrepanz zwischen tatsächlichem und angemessenem Vertrauen beurteilen lässt.

5.1.2 Vertrauen als Prozess der Auflösung multipler Ambiguitäten

Die qualitativen Aussagen der Versuchsteilnehmenden in den durchgeführten Studien und deren hohe Sensitivität für die Beschreibungen der Cobots legen die Vermutung nahe, dass diese die Cobots als ambig, also als mehrdeutig wahrnahmen¹. Der Begriff der Ambiguität wird kontextspezifisch unterschiedlich verwendet und bedarf daher einer näheren Bestimmung. Für den MRI-Kontext scheint die Definition nach Caglioti (1992, S. 17) geeignet, die die „Koexistenz zweier sich gegenseitig ausschließender Aspekte oder Schemata der Realität“ (zitiert nach vgl. M. Bauer, Knape, Koch & Winkler, 2010, S. 12) als Entstehungsbedingung für Ambiguität herausstellt². Dabei ist zu beachten, dass zwar Situationen oder Objekte ambige Stimuli darstellen können, Ambiguität allerdings letztlich erst im Rahmen des individuellen Wahrnehmungsvorgangs entsteht und somit stets ein subjektives Moment beinhaltet. Somit kann die bloße Wahrnehmung von „Inkonsistenzen oder Widersprüchlichkeiten vorliegender Informationen“ (Ziegler, 2010, S. 126) zu einer

¹ Vgl. hierzu die Ergebnisse der qualitativen Vorstudie 2 in Kapitel 4.3.3 sowie die qualitativen Auswertungen zu den Experimentalstudien in Kapitel 4.5.3.9 und Kapitel 4.6.3.7. Besonders plakativ formulierte bspw. ein Versuchsteilnehmer in Experimentalstudie 1, dass es sehr schwer vorherzusehen sei, ob der Cobot ihm „den Arbeitsplatz in Zukunft streitig machen würde oder nicht“.

² Alternative Definitionen von Ambiguität entstammen bspw. der Entscheidungstheorie oder der Linguistik. In der Entscheidungstheorie bezieht sich der Begriff auf die durch Informationsmangel bedingte Unsicherheit über die Eintrittswahrscheinlichkeit von Handlungsalternativen (Stocké, 2002, S. 37f.). Im Bereich der Textlinguistik und kognitiven Linguistik werden verschiedene Formen sprachlicher Ambiguität in Zusammenhang mit Frames diskutiert (M. Bauer, Knape, Koch & Winkler, 2010, S. 10).

(wahrgenommenen) Ambiguität führen. Die Mehrdeutigkeit bezieht sich dabei – in Abgrenzung zum Begriff der Ambivalenz – explizit auf die Wahrnehmung anstatt auf die Bewertung des Betrachtungsgegenstandes (Ziegler, 2010, S. 126f.).

In Bezug auf Cobots lassen sich solche Mehrdeutigkeiten in Form von konfligierenden Stimuli und mentalen Modellen auf verschiedenen Ebenen herausarbeiten, wie Abbildung 5.2 visualisiert. Zunächst erscheint weder der ontologische Status des Cobots noch die Mensch-Cobot-Relation als eindeutig. So repräsentieren Mitarbeitende Cobots als menschenähnlich oder maschinell und die Relation zu ihnen als kooperierend oder konkurrierend. Die Ambiguität manifestiert sich dabei sowohl in Bottom-Up- als auch in Top-Down-Verarbeitungsprozessen. Auf der Ebene von Top-Down-Schemata führen koexistierende technikhorientierte Betrachtungen von Cobots, bspw. durch die Darstellung technologischer Parameter wie der Tragkraft und der Bewegungsachsen, und die stark vermenschlichende Sprache, die mitunter Zuschreibungen von Intelligenz und Autonomie beinhaltet, zu einer Unsicherheit bezüglich des ontologischen Status (vgl. Kapitel 2.2.4 zu hybriden Robotern). Auf Ebene der Bottom-Up-Reize steht die grundsätzlich funktionale Gestaltung im Widerspruch zu mitunter implementierten anthropomorphen Designmerkmalen und sozialen Funktionalitäten, die gar den Mehrwert im funktionalen Anwendungskontext unterminieren können (Roesler & Onnasch, 2020a). Ein vergleichbarer Widerspruch ergibt sich überdies aus der Kollaborationsfähigkeit von Cobots auf der einen Seite, die in Form direkter physischer Interaktion beobachtbar wird und eine aktive Beteiligung der Mitarbeitenden am Produktionsprozess nahelegt, und der häufig koexistenten Anwendungsszenarien in der Praxis auf der anderen Seite, die die Mitarbeitenden aus der Arbeitstätigkeit ausschließen und damit ein Konkurrenzverhältnis konstruieren (vgl. Kapitel 2.2.3 und 2.3.2 zu realen Anwendungsszenarien von Cobots). Als gegensätzliche Top-Down-Einflüsse auf die Wahrnehmung der Mensch-Roboter-Relation sind ferner die konkurrierenden und konfligierenden Narrative aus der öffentlichen Debatte zu nennen, die abweichende anthropologische Selbstdeutungen in Bezug auf das Mensch-Technik-Verhältnis implizieren (Heßler, 2019). Neben dem Narrativ der unterstützenden Assistenztechnologie kursieren Umschreibungen von Robotern als Job-Killer, die einer Rationalisierungslogik folgen. In Summe ergibt sich eine Vielzahl an unterschiedlich manifestierten Mehrdeutigkeiten und damit verbundenen Irritationen in der Betrachtung von Cobots, die sich u. a. in den viel diskutierten Dichotomien zwischen Mensch und Maschine sowie zwischen Werkzeug und Teammitglied niederschlagen (Groom & Nass, 2007; Roesler & Onnasch, 2020b). In den durchgeführten Experimentalstudien standen dabei die existierenden Ambiguitäten auf Top-Down-Verarbeitungsebene sowie deren Disambiguierung durch Framing im Fokus.

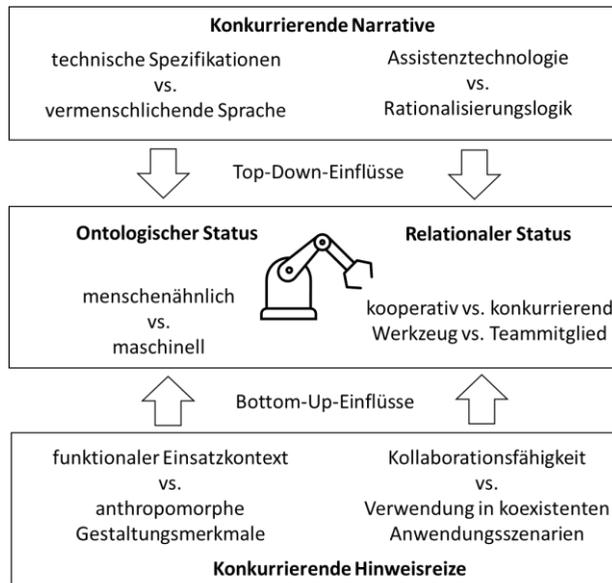


Abbildung 5.2: Schaubild über multiple Ambiguitäten von Cobots, die sich auf dessen ontologischen und relationalen Status beziehen und sich in Top-Down-Schemata und Bottom-Up-Reizen manifestieren.

Die ambige Wahrnehmung von Cobots spiegelt sich auch in früheren Studienergebnissen wider (vgl. auch Kapitel 2.2.4 zu hybriden Robotern). Diese zeigten, dass Personen Roboter mitunter als weder *lebendig* noch *unlebendig* und weder *geföhlsbegabt* noch *nicht-geföhlsbegabt* umschreiben (Damiano & Dumouchel, 2018, S. 4), deren Aktionen häufig sehr genau beobachten und zu ambivalenten Bewertungen derselben neigen (Law, Malle & Scheutz, 2021). In einer aktuellen Publikation beschreiben Matthews et al. (2021, S. 60) moderne Roboter (und ähnliche Technologien) gar als „fundamental ambig[e]“ (*fundamentally ambiguous*) Systeme, die sich mental nur schwer eindeutig in bekannten Schemata repräsentieren lassen, sondern einen Bereich zwischen hochgradig fortschrittlichen Maschinen und Menschen besetzen.

Die Erfahrung von Ambiguität ist mit Geföhlen der Unsicherheit und einer hohen kognitiven Beanspruchung verknüpft (Stocké, 2002, S. 37f.) und konstituiert damit die Notwendigkeit der Vertrauensbildung. Die Unfähigkeit, das jeweilige Gegenüber in gedanklichen Kategorien zu verorten, bringt die Unsicherheit darüber mit sich, wie dieses Gegenüber zukünftig agieren wird. Verschiedenartige Zukunftserwartungen und -entwürfe müssen daher parallel bedacht werden. Diese kognitive Beanspruchung mindert die zur Verfügung stehende kognitive Kapazität. Entsprechend führte in früheren Experimenten die Präsentation konfligierender Stimuli zu einer schlechteren Performance eines Mensch-Roboter-

Teams (Smith & Wiese, 2016). Es verwundert daher wenig, dass Menschen Ambiguität tendenziell aversiv gegenüberstehen (Ziegler, 2010, S. 150), wie sich anhand des sog. *Ellsberg-Paradoxons* aus der Entscheidungstheorie (Ellsberg, 1961) zeigen lässt. Aus der wahrgenommenen Ambiguität und der damit verbundenen aversiven Reaktion entspringt eine starke Motivation, sich auf ein eindeutiges mentales Modell der gegenwärtigen Situation und durch dessen mentale Simulation auf einen bestimmten Zukunftsentwurf festzulegen. In Anlehnung an Luhmann (1968/2014) lässt sich genau diese komplexitätsreduzierende Festlegung als Akt des Vertrauens (oder des Misstrauens) verstehen (vgl. Kapitel 3.4.1). Letztlich vertrauen die jeweiligen Akteur:innen primär darauf, dass es sich bei einer bestimmten Repräsentation um ein zweckdienliches *Modell für* die Antizipation künftiger Vorgänge in der Umwelt handelt (vgl. Kapitel 3.1.1). Der Akt des Vertrauens stellt damit einen Mechanismus zur mentalen Disambiguierung einer als ambig wahrgenommenen Situation oder Konstellation dar. Vor diesem Hintergrund lässt sich auch die in den qualitativen Studien beobachtete hohe Aufmerksamkeit und Sensitivität in Hinblick auf die vermittelten Informationen über einen Cobot deuten.

5.1.3 Vertrauen am Arbeitsplatz als multidimensionales Phänomen

Bereits in Vorstudie 2 verdeutlichten die Aussagen der Produktionsmitarbeitenden, dass sich das Vertrauen in einen Cobot nicht eindimensional beurteilen lässt (vgl. Kapitel 4.3.3), sondern weitere Differenzierungen des jeweiligen Bezugsobjektes des Vertrauens erforderlich sind. Auch die gemessenen Auswirkungen der wahrgenommenen Menschenähnlichkeit und der Mensch-Cobot-Relation auf das Vertrauen in Experimentalstudie 1 ließen sich nur unter Einbezug verschiedenartiger Ziele der Mitarbeitenden und der daraus resultierenden Bezugsobjekte des Vertrauens verstehen (vgl. Kapitel 4.5.4.2). Das steht in Einklang mit der jüngeren MRI-Forschung, die zunehmend die Komplexität, den Facettenreichtum und die Multidimensionalität von Vertrauen offenlegt. Das gilt auch (oder gerade) für das Mensch-Roboter-Vertrauen in anwendungsnahen und funktionalen Nutzungskontexten (A. L. Baker, Phillips, Ullman & Keebler, 2018; Schaefer et al., 2021, S. 281; Ullman & Malle, 2018). Bei der empirischen Erhebung eines Vertrauensniveaus ist dabei stets zu hinterfragen, mit welchen Erwartungen dieses Vertrauen einhergeht, welche Vertrauenskomponenten hierbei eine Rolle spielen und was das Bezugsobjekt des Vertrauens darstellt. In Kapitel 3.4.4 wurde bereits auf die Differenz zwischen kognitiven und affektiven Vertrauenskomponenten hingewiesen. Die voneinander verschiedenen Vertrauensverläufe in beiden Experimentalstudien lassen sich neben populationspezifischen Einflussfaktoren (vgl. Kapitel 4.7.1) alternativ auch durch diesen Unterschied erklären. Online-Studien vermögen dabei nur ein begrenztes Bild des ganzheitlichen Vertrauens abzubilden,

da in diesem Setting keine vergleichbaren affektiven Reaktionen auftreten wie bei der tatsächlichen Begegnung mit einem Roboter in einem Präsenzsetting.³

Eine weitere Unterscheidungsdimension bei der Analyse des Mensch-Cobot-Vertrauens stellt ferner der Kontext dar, in dem eine MRI stattfindet. Dieser lässt sich bspw. nach Lewis et al. (2018) als (primär) leistungsorientierter (*performance-based*) oder sozial-orientierter (*social-based*) Interaktionskontext klassifizieren. Analog unterteilen Jessup und Schneider (2021, S. 523) die verfügbaren Roboter-Typen in Performance-Roboter und soziale Roboter. Daraus lässt sich allerdings weder schließen, dass eine dichotome Entweder-oder-Differenz zwischen den Interaktionskontexten besteht, noch dass sich gewisse Robotertypen diesen Kontexten eindeutig zuordnen lassen (vgl. die Argumentation zu hybriden Robotern in Kapitel 2.2.4). Die Ergebnisse der Vorstudien lassen sich derart interpretieren, dass sowohl leistungsorientierte als auch sozial-orientierte Faktoren das Vertrauen am Arbeitsplatz beeinflussen. Zwar stellen soziale Interaktionen in Anwendungsbereichen wie der industriellen Produktion keinen Selbstzweck dar, können aber dennoch einen wichtigen vertrauensbildenden Faktor und ein Mittel zum Zweck der performanten Erfüllung von Arbeitsaufgaben darstellen. Insofern unterscheiden sich leistungs- und sozial-orientierte Anwendungskontexte lediglich dahingehend, dass soziale Interaktionen entweder einen Selbstzweck oder ein Mittel zum Zweck der Erfüllung leistungsorientierter Zielkriterien darstellen. Umso mehr ist die Frage nach dem Vertrauen im industriellen Kontext als vielschichtig zu betrachten und kann verschiedene Bezugsobjekte betreffen.

Im Rahmen der Entwicklung eines multidimensionalen Vertrauensmodells identifizierten Ullman & Malle (2018) kürzlich eine weitere Unterscheidungsdimension für Vertrauenskomponenten, auf die sich die Ergebnisse dieser Dissertation beziehen lassen. Auf Basis einer empirischen Analyse des „semantischen Raumes von Vertrauen“ (*semantic space of trust*) (Ullman & Malle, 2018, S. 263) unterscheiden die Autoren sog. leistungsorientiertes (*performance trust*) und moralisches Vertrauen (*moral trust*) (Malle & Ullman, 2021; Ullman & Malle, 2018). Analog ordnen Law und Scheutz (2021) im Rahmen einer Überblicksstudie die Vertrauenskonzepte in bisherigen empirischen Studien entweder dem leistungsorientierten (*performance-based*) oder dem relationsorientierten (*relation-based*) Vertrauen zu. Sie verorten leistungsorientiertes Vertrauen ebenfalls im Kontext industrieller MRI und relationsorientiertes Vertrauen im klassischen Anwendungskontext der sozialen Robotik. Gleichwohl stellen sie eine „weit verbreitete Verquickung“ (*fairly widespread conflation*) (Law & Scheutz, 2021, S. 50) beider Vertrauenskonzepte fest.

Leistungsorientiertes Vertrauen bezieht die Fähigkeiten (*capable*) und die Zuverlässigkeit (*reliable*) der vertrauensnehmenden Instanz ein, wohingegen das moralische bzw.

³ Eine ausführliche Darstellung des Einflusses von Emotionen und Affekten auf die Wahrnehmung einer MRI findet sich in Jessup und Schneider (2021).

relationsorientierte⁴ Vertrauen durch ethisches Verhalten (*ethical*) und Aufrichtigkeit (*sincere*) bestimmt wird (Ullman & Malle, 2019). Letzteres basiert auf der Erwartung, dass die vertrauensnehmende Instanz in gewisser Weise einen Teil der Gesellschaft darstellt, moralische Verhaltensweisen an den Tag legt und bestehende Vulnerabilitäten der vertrauenden Person nicht ausnutzen wird (Law & Scheutz, 2021; Malle & Ullman, 2021). Diese Unterscheidung erinnert an die bereits thematisierten Kriterien für die Vertrauenswürdigkeit von Vertrauensnehmer:innen (vgl. Kapitel 3.4.1). Diese umfassen die Fähigkeiten, die wohlwollende Haltung und die Integrität der vertrauensnehmenden Instanz (Mayer et al., 1995), wobei die beiden letzten Kriterien Bestandteile des relationsorientierten Vertrauens darstellen (Malle & Ullman, 2021). Insofern besteht der Neuigkeitswert des multidimensionalen Vertrauensmodells weniger in der Identifikation unterschiedlicher Teilaspekte des Vertrauens, sondern in der These, dass diese unterschiedlichen wahrgenommenen Merkmale der vertrauensnehmenden Instanz nicht in ein kollektives Vertrauensempfinden münden, sondern gewisse Arten des Vertrauens koexistieren können (vgl. hierzu auch die Koexistenz von Vertrauen und Misstrauen in Kapitel 3.4.10). Tabelle 37 stellt die verschiedenen Dimensionen von Vertrauen sowie die möglichen Ausprägungen dar und demonstriert damit die Multidimensionalität von Vertrauen. Die Unterscheidung zwischen relations- und leistungsorientiertem Vertrauen wird in Hinblick auf die Zusammenhänge zwischen Vertrauen und Misstrauen im folgenden Kapitel 5.1.4 sowie zur Darstellung der multidimensionalen Auswirkungen des Anthropomorphismus in Kapitel 5.3.2 erneut aufgegriffen.

Tabelle 37: Dimensionen und mögliche Ausprägungen von Vertrauen.

	Mögliche Ausprägungen der Dimension			Quelle
Zeitliche Komponente	initial	dynamisch		Hoff und Bashir (2015)
Verarbeitungsmodus	kognitiv	affektiv		Lewicki et al. (2006)
Entstehungslogik	analytisch	analogisch	affektiv	J. D. Lee und See (2004)

⁴ Im Folgenden wird der Begriff des relationsorientierten Vertrauens verwendet, da ein Vertrauen in das moralische Handeln einer Maschine insofern schwer vorstellbar erscheint, als Maschinen gemeinhin nicht als moralische Akteur:innen mit eigenem Moralverständnis begriffen werden. Allenfalls ließe sich aus einer ausschließlich auf die Handlungsfolgen fokussierten Sicht darauf vertrauen, dass ein Roboter Aktionen ausführt, deren Folgen eine vertrauende Person vor dem Hintergrund ihres jeweiligen Moralverständnisses als moralisch gut bewertet. Aus dieser Sichtweise handelte es sich bei einem solchen Roboter um eine „einfache moralische Maschine“ (Bendel, 2019a, S. 40).

	Mögliche Ausprägungen der Dimension			Quelle
	mensch-orientiert	roboter-orientiert	kontextuell	
Arten von Einflussfaktoren	mensch-orientiert	roboter-orientiert	kontextuell	Hancock, Billings und Schaefer, et al. (2011)
Arten von Bezugsobjekten des Vertrauens	leistungsorientiert	moralisch / relationsorientiert		Malle und Ullman (2021) bzw. Law und Scheutz (2021)
Kriterien für Vertrauenswürdigkeit	Fähigkeiten	Wohllwollen	Integrität	Mayer et al. (1995)

5.1.4 Vertrauen und Misstrauen als unterschiedliche Konstrukte

Im Theorieteil wurde bereits die Frage thematisiert, ob es sich bei Vertrauen und Misstrauen um die Endpunkte einer gemeinsamen Skala oder um disjunkte Konstrukte handelt (vgl. Kapitel 3.4.10). Wie Abbildung 5.3 und Tabelle 38 zeigen, entwickelten sich Vertrauen und Misstrauen in den Experimentalstudien zumeist antiproportional, was zumindest auf einen starken Zusammenhang beider Konstrukte hindeutet. Gleichwohl legen die im Falle des Framing-Effekts in Experimentalstudie 2 stark unterschiedlichen Muster (vgl. Kapitel 4.7.3) die Vermutung nahe, dass Vertrauen und Misstrauen von unterschiedlichen Wirkmechanismen beeinflusst werden. Diese Schlussfolgerung wird zusätzlich dadurch gestützt, dass stärkere Bedenken und Befürchtungen wie Jobverlust, Verletzungsrisiko und soziale Isolation signifikant mit dem Misstrauensniveau korrelierten. Das Vertrauensniveau stand hingegen mit positiven Aussagen in Zusammenhang, die den Mehrwert durch die Robotereinführung adressieren, wie z. B. Spaß bei der Benutzung und eine hohe Nützlichkeit (vgl. Abbildung 4.35 in Kapitel 4.6.4.3). Ferner zeigte sich in Experimentalstudie 1, dass Misstrauen weniger stark durch die wahrgenommene Menschenähnlichkeit und die wahrgenommene Kooperationsorientierung der Mensch-Cobot-Relation beeinflusst wird als Vertrauen. Einen hypothetischen Erklärungsansatz hierfür liefert der bereits in Kapitel 5.1.1 thematisierte Übereinstimmungsbias aus der Sozialpsychologie, demzufolge negative Einstellungen veränderungsresistenter sind als positive.

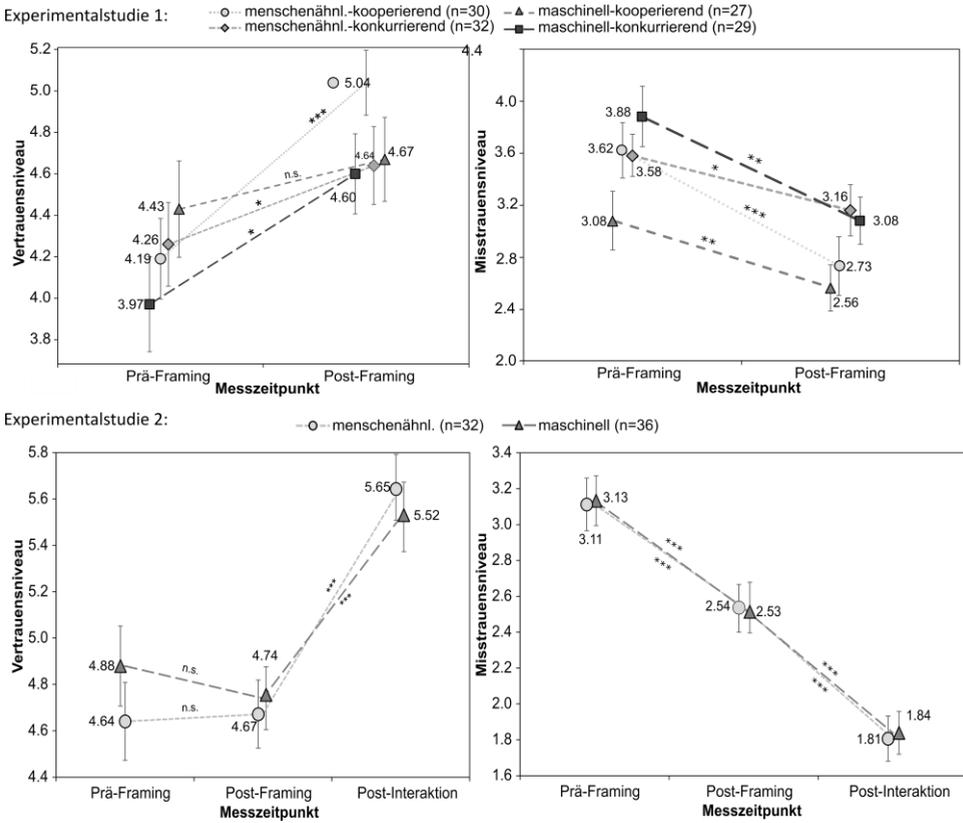


Abbildung 5.3: Gegenüberstellung der jeweiligen Verläufe von Vertrauen und Misstrauen in den beiden Experimentalstudien.

Tabelle 38: Korrelationen zwischen Vertrauen und Misstrauen zu verschiedenen Messzeitpunkten in beiden Experimentalstudien.

	vor dem Framing		nach dem Framing		nach der Interaktion	
	<i>r</i>	<i>p</i>	<i>r</i>	<i>p</i>	<i>r</i>	<i>p</i>
Experimentalstudie 1	-.508***	< .001	-.535***	< .001	-	-
Experimentalstudie 2	-.497***	< .001	-.345**	.004	-.138	.261

Gleichwohl besteht die Möglichkeit, dass die beobachteten partiell unterschiedlichen Verläufe von Vertrauen und Misstrauen in Wirklichkeit den Unterschied zwischen leistungs- und relationsorientiertem Vertrauen abbilden. Ordnet man die einzelnen Items der verwendeten TAS diesen beiden Vertrauensdimensionen zu (Malle & Ullman, 2021), fällt auf, dass drei der vier zugeordneten Misstrauensitems, aber nur eins der vier zugeordneten

Vertrauensitems relationsorientiertes Vertrauen charakterisieren (vgl. Tabelle 39). Die Einteilung in vertrauens- und misstrauensbezogene Items überschneidet sich demnach stark mit der Einteilung in leistungs- und relationsorientierte Vertrauenskomponenten. Die vermutete Differenz zwischen Vertrauen und Misstrauen konfundiert daher mit derjenigen zwischen relations- und leistungsorientiertem Vertrauen. Alternativ zur Vermutung aus Kapitel 4.7.3, dass sich imaginierter Kontakt stärker auf Misstrauen auswirkt, ließe sich auf Basis dieses konfundierenden Zusammenhangs annehmen, dass imaginierter Kontakt stärker auf relationsorientierte Vertrauenskomponenten wirkt. Weitere empirische Studien sind zur Abklärung erforderlich, auf welchen konzeptionellen Unterschied sich die beobachteten empirischen Verlaufsmuster zurückführen lassen.

Tabelle 39: Items der TAS mit Zuordnung zu den Dimensionen des Vertrauens, die damit gemessen werden.

Item	Konstrukt	Zuordnung nach Malle und Ullman (2021)
1 Der Roboter ist irreführend.	Misstrauen	Relationsorientiertes Vertrauen
2 Der Roboter verhält sich undurchsichtig.	Misstrauen	Relationsorientiertes Vertrauen
3 Ich misstrauere den Entscheidungen des Roboters.	Misstrauen	Relationsorientiertes Vertrauen
4 Ich muss vorsichtig im Umgang mit dem Roboter sein.	Misstrauen	<i>nicht zugeordnet</i>
5 Die Handlungen des Roboters haben negative Auswirkungen zur Folge.	Misstrauen	Leistungsorientiertes Vertrauen
7 ⁵ Der Roboter bietet Sicherheit.	Vertrauen	Leistungsorientiertes Vertrauen
8 Der Roboter arbeitet tadellos.	Vertrauen	Relationsorientiertes Vertrauen ⁶
9 Der Roboter ist verlässlich.	Vertrauen	Leistungsorientiertes Vertrauen
10 Der Roboter ist vertrauenswürdig	Vertrauen	Leistungsorientiertes Vertrauen
11 Ich kann dem Roboter vertrauen.	Vertrauen	<i>nicht zugeordnet</i>

⁵ Ein sechstes Item ist in der Tabelle nicht enthalten, da das sechste Item der Originalskala nicht in die deutsche Skala übernommen wurde.

⁶ Das englischsprachige Item *The system has integrity* weist einen deutlich klareren Bezug zum relationsorientierten Vertrauen auf. Auf Basis der deutschen Formulierung erscheint eine Zuordnung zum leistungsorientierten Vertrauen ebenfalls als möglich.

Item	Konstrukt	Zuordnung nach Malle und Ullman (2021)
12 Ich kenne mich mit dem Roboter aus.	Vertrauen	<i>nicht zugeordnet</i>

5.2 Mensch-Roboter-Relationen und die Angst vor Arbeitsplatzverlust

5.2.1 Roboter als (vermeintliche) Teamkolleg:innen

Seit Groom und Nass (2007) schlussfolgerten, dass Roboter keine Teamkolleg:innen sind, haben Forschung und Entwicklung im Roboterbereich große Fortschritte erzielt. Inzwischen weist der Tenor in der Forschung eher in die Richtung, dass Roboter zumindest als Teamkolleg:innen wahrgenommen werden können (Matthews et al., 2021, S. 60). So kommen Nyholm und Smids (2020) in einer ausführlichen Analyse zu dem Schluss, dass moderne Roboter imstande sind, die meisten Kriterien für Kolleg:innen zu erfüllen. Sie begründen dies damit, dass die Beurteilung von Kolleg:innen kontextabhängig und domänenspezifisch erfolge. Im Gegensatz zu Freund:innen oder Liebespartner:innen ist das tatsächliche Innenleben von Kolleg:innen weitestgehend irrelevant. Gleichwohl vermag nicht jeder Roboter, sämtliche mit Kolleg:innen assoziierten typischen Ideale zu erfüllen. Roboter könnten bspw. im Vergleich zu menschlichen Kolleg:innen eine bessere Unterstützung bei der Bewältigung von Arbeitsaufgaben und zugleich schlechtere Konversationspartner:innen darstellen (Nyholm & Smids, 2020). Smids, Nyholm und Berkers (2020) argumentieren sogar, dass Roboter am Arbeitsplatz so gestaltet werden *sollten*, dass die Mitarbeitenden sie als Teamkolleg:innen betrachten. Sie sehen darin die Chance für eine stärker sinnstiftende Arbeit, wenn die Mitarbeitenden ein Gefühl der Zugehörigkeit zu den Robotern entwickeln und damit Bedenken bezüglich einer zunehmenden sozialen Isolation ablegen. Allerdings lässt sich in Frage stellen, ob solche quasi-sozialen Beziehungen die entsprechenden Bedürfnisse der Mitarbeitenden tatsächlich zu befriedigen imstande sind, insofern der betreffende Roboter Gefühle gegenüber dem Mitarbeitenden allenfalls simulieren und damit eine bloße Illusion einer kollegialen Beziehung erzeugen kann (vgl. dazu Turkle, 2010).

Inzwischen existieren erste breiter angelegte empirische Studien zur Wahrnehmung von Mensch-Roboter-Relationen. Beispielsweise verdeutlichte eine qualitative Untersuchung in den USA, dass ca. 30 % von $n = 605$ befragten Arbeiter:innen ihre Beziehung zu einer beliebigen intelligenten Technologie als kollegial (*teammate-like based*) beschrieben

(Lyons, Wynne, Mahoney & Roebke, 2019). Da die Autoren den Begriff der intelligenten Technologie dabei sehr weit fassten, sodass dieser ein Spektrum von autonomen Fahrzeugen über Sprachassistenten bis hin zu industriellen Robotern abdeckte, bleibt die Übertragbarkeit auf MRI-Situationen jedoch unklar. Die Ergebnisse der Manipulationskontrolle aus der ersten Experimentalstudie (vgl. Kapitel 4.5.3.1) liefern zumindest eine erste empirische Unterstützung für die These, dass Menschen kooperationsorientierte Relationen zu Cobots konstruieren. In Anlehnung an die Theorie des dualen Anthropomorphismus (vgl. Kapitel 3.2.3.2) ist ferner davon auszugehen, dass explizite Messmethoden ein verzerrtes Bild liefern. Einige Personen könnten auf explizite Nachfrage hin verneinen, dass sie einen Roboter als Teammitglied betrachten, da sie davon ausgehen, dass dies der sozialen Norm zuwiderläuft. Gleichwohl besteht die Möglichkeit, dass sie sich – analog zu Prozessen der schwachen Vermenschlichung (vgl. Kapitel 3.2.2) – einem Roboter gegenüber so verhalten, *als ob* sie ihn als Teammitglied betrachten (vgl. hierzu auch die *in situ*-Studien in Kapitel 2.3.2.2).

Die Beurteilung, inwiefern aus Robotern Teamkolleg:innen werden können, hängt sehr stark von den Ansprüchen an Teamkolleg:innen ab. Modelle zur wahrgenommenen Kollegialität entstanden erst in den letzten Jahren und beziehen sich häufig nicht auf eine spezielle Technologie, sondern auf allgemeiner gefasste Klassen wie bspw. auf autonome Akteur:innen. Wynne und Lyons (2018, S. 355) liefern eine ambitionierte Definition für Teamkolleg:innen als „hochgradig altruistisch, wohlwollend, voneinander abhängig, gefühlsbetont, kommunikativ und synchronisiert agierend“ (Übersetzung des Verfassers). Damit lassen sich bereits erste Einflussfaktoren ableiten, die die Wahrnehmung als Teamkolleg:in begünstigen. Gemeinhin zählen insbesondere anthropomorphe Merkmale dazu (W. Kim et al., 2020; Lyons et al., 2019; Matthews et al., 2021, S. 60; Wynne & Lyons, 2018, S. 366). Analog zu den Vertrauensmodellen ist allerdings davon auszugehen, dass überdies auch menschenbezogene Merkmale wie der Charakterzug der Verträglichkeit oder die Tendenz zur Objektifizierung (Wynne & Lyons, 2018, S. 364) und situative Faktoren wie die aufgabenbezogene Abhängigkeit vom jeweiligen Gegenüber sowie die Aktivierung bestehender mentaler Modelle die Wahrnehmung eines Roboters als Teamkolleg:in beeinflussen. Diese wird wiederum mit positiven Konsequenzen für die MRI assoziiert. Forschende gehen davon aus, dass diese Wahrnehmung die Zusammenarbeit vereinfacht, die gemeinsame Leistungsfähigkeit und die Fehlertoleranz erhöht, insbesondere wenn eine hohe Unsicherheit hinsichtlich der allgemeinen Technologieverwendung besteht (Alarcon et al., 2021, S. 173; Matthews et al., 2021, S. 68f.; Wynne & Lyons, 2018, S. 368). Insofern sind vertrauenssteigernde Effekte solcher Relationen insbesondere in Hinblick auf relationsorientiertes Vertrauen anzunehmen. Diese könnten sogar stärker und eindeutiger ausfallen als die intensiv beforschten Effekte des Anthropomorphismus, die – wie u. a. in Experimentalstudie 1 gezeigt – kontextspezifisch und je in Bezug zu den verschiedenen Vertrauensdimensionen unterschiedlich ausfallen (vgl. Kapitel 5.3.2). Eine vertiefte Untersuchung der Zusammenhänge zwischen der wahrgenommenen

Menschenähnlichkeit und der konstruierten Mensch-Cobot-Relation innerhalb verschiedener Kontexte erscheint vor diesem Hintergrund besonders aussichtsreich.

Relationale Faktoren werden bis dato kaum in Vertrauensmodellen repräsentiert und empirisch berücksichtigt. Diese Modelle differenzieren häufig zwischen menschenorientierten, roboterbezogenen und situativen oder kontextuellen Faktoren (vgl. Hancock, Billings, Schaefer et al., 2011). In Bezug auf Cobots bleibt dabei unberücksichtigt, dass angesichts der neuen Interaktionsmöglichkeiten die beteiligten Personen ein gemeinsames sozio-technisches Arbeitssystem mit dem Cobot bilden, das von relationalen Aspekten wie den Mechanismen der Zusammenarbeit geprägt ist. Nachdem lange Zeit roboterbezogene Faktoren die Forschung dominierten, erhielten in den vergangenen Jahren die personenbezogenen Faktoren mehr Aufmerksamkeit. Bemerkenswerterweise berücksichtigen allerdings nur sehr wenige Untersuchungen, wie beteiligte Personen die Relation zu dem entsprechenden Roboter wahrnehmen, d. h. auf welche Art und Weise sich diese Personen zu einem bestimmten Roboter in Beziehung setzen (Wynne & Lyons, 2018, 2019). Womöglich lässt sich das darauf zurückführen, dass frühere Forschung trivialisierend von einer hierarchischen Relation zwischen Menschen als Werkzeugverwendende und Roboter als Werkzeuge ausging. Inzwischen wächst das Bewusstsein dafür, dass die klassische Werkzeugverwender-Werkzeug-Relation eine Verkürzung darstellt, die sich kaum mehr auf moderne Mensch-Roboter-Relationen anwenden lässt, insbesondere angesichts hybrider Roboter und neuer Interaktionsformen. Zwar scheint es – zumindest zum aktuellen Stand der Technikentwicklung – ausgeschlossen, dass Roboter im industriellen Umfeld in gleichem Ausmaß zu Teamkolleg:innen werden wie Menschen. Diese Feststellung impliziert allerdings nicht, dass sich Menschen in diesem Kontext überhaupt nicht oder nur funktional-hierarchisch zu einem Roboter in Beziehung setzen (Lyons & Wynne, 2021). Die empirischen Erkenntnisse in der vorliegenden Dissertation unterstreichen den Variantenreichtum und die Relevanz der konstruierten Relationen und werfen damit ein Schlaglicht auf eine Facette industrieller MRI, die bisher insbesondere in empirischen Studien ein Schattendasein führte. Daher scheint es sinnvoll, künftig die kontextuellen Faktoren unter der Perspektive eines sozio-technischen Arbeitssystems und mit einem verstärkten Fokus auf die relationalen Aspekte wissenschaftlich zu untersuchen.

5.2.2 Relationale Wende in der MRI

Die zuvor skizzierten Lücken der bisherigen anwendungsnahen empirischen Forschung lassen sich vor dem Hintergrund technikphilosophischer Strömungen kontextualisieren. Ursprünglichen wurden die Debatten um die Mensch-Maschine-Relation vorwiegend auf einen Dualismus zwischen Teamkolleg:in und Werkzeug zugespitzt. Seit kurzem mehren sich Forderungen einiger Philosoph:innen, die Betrachtungsweise von Robotern als reine

Werkzeuge grundsätzlich zu überdenken (Smids et al., 2020, S. 16) und damit das Feld für eine differenziertere Betrachtung zu öffnen.

Anknüpfend an die früheren Veröffentlichungen von Coeckelbergh (2010b, 2011) erscheinen phänomenologische Betrachtungen zunehmend aussagekräftiger als ontologische Fragestellungen, um die Reaktionen und Relationen von Menschen im MRI-Kontext besser zu verstehen. Die entsprechende Fragestellung zielt dann darauf ab, wie die Roboter den beteiligten Personen *erscheinen* und nicht darauf, was sie *sind* (Wynne & Lyons, 2018, S. 360). Dieser phänomenologische Blickwinkel kann sich sowohl auf die Erscheinung des Cobots selbst als auch auf die Relation zu demselben beziehen. Die Erläuterungen in Kapitel 5.2.1 haben die These begründet, dass Roboter aus einer phänomenologisch-relationalen Perspektive durchaus als Teamkolleg:innen – oder neutraler formuliert: als Teammitglieder – wahrgenommen werden können. Diese Debatte berührt philosophische Fragen, die insbesondere von Roboterethiker:innen und Technikphilosoph:innen verstärkt diskutiert werden. So prägten und entwickelten Coeckelbergh (2010a, 2012) sowie Coeckelbergh und Gunkel (2014) den Begriff der „relationalen Wende“ (*relational turn*) (Gunkel, 2018, S. 96) in der Ethik. Dieser rekurriert auf eine Verschiebung der relevanten Kriterien zur Bestimmung des moralischen Status einer Entität. Dementsprechend stellen nicht die Eigenschaften eines Roboters, sondern dessen Situierung in Relation zum jeweiligen Gegenüber die Grundlage für die Festlegung des sozialen und moralischen Status dieses Roboters dar. Die Wurzeln für diesen Paradigmenwechsel sieht Gunkel (2018) in Emmanuel Lévinas' These (vgl. Lévinas & Krewani, 1993), wonach Menschen moralische Betrachtungen nicht auf der Basis eines ontologischen Status, sondern vorgelagert „im Angesicht aktueller sozialer Beziehungen und Interaktionen“ (*in the face of actual social relationships and interactions*) (Gunkel, 2018, S. 96) vornehmen. Ferner verteidigt bspw. Danaher (2020, S. 2027) im Rahmen des ethischen Behaviorismus die These, dass sich moralische Verpflichtungen gegenüber Robotern aus den „beobachtbaren behavioralen Relationen“ (*observable behavioural relations*) ableiten. Gemein ist diesen inklusiven Ansätzen der Roboterethik (vgl. dazu Loh, 2019, S. 95ff.) eine kritische Haltung gegenüber dem tradierten Subjekt-Objekt-Dualismus und eine Verlagerung des Denkens weg „von den Relata (...) zu den Relationen, den eigentlichen Praktiken sowie dem sozialen Kontext“ (Loh, 2019, S. 115). Auch Turkles (2002, S. 150) Umschreibung von Robotern als „relationale Artefakte“ (*relational artifacts*) verweist bereits explizit auf die relationale Ebene.

Abbildung 5.4 bietet eine Einordnung möglicher Fragestellungen im Kontext der industriellen MRI anhand philosophischer Kategorien von ontologischen über dichotome phänomenologische bis hin zu weiter ausdifferenzierten Fragestellungen mit einem Fokus auf relationalen Aspekten. Die Abfolge der Fragestellungen von oben nach unten spiegelt eine zunehmende Fokussierung auf diejenigen Aspekte wider, die entscheidend sind, um die menschlichen Reaktionsmuster in Bezug auf Cobots zu verstehen.

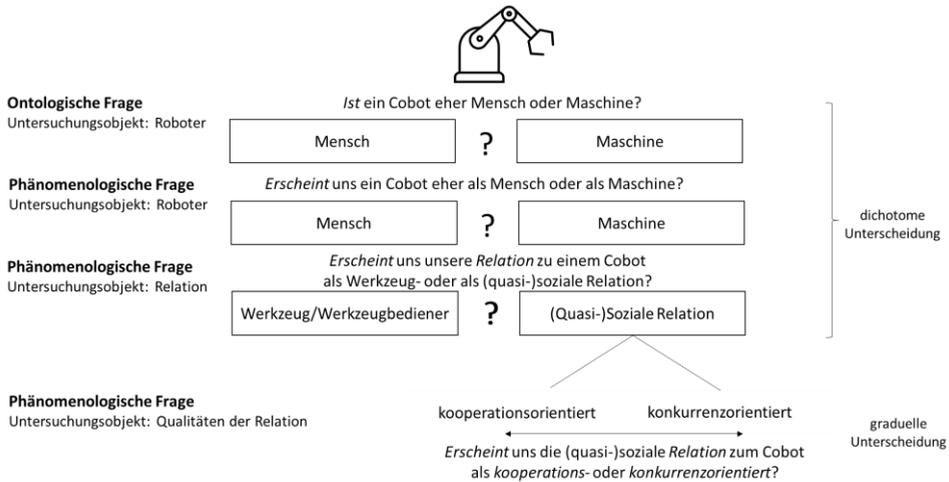


Abbildung 5.4: Philosophische Einordnung möglicher Fragestellungen und Forschungsfokusse im industriellen MRI-Kontext.

Während die traditionelle Unterscheidung zwischen Menschen und Maschinen gemeinhin als kategorial betrachtet wird, erlauben die relationsbezogenen Differenzierungen graduelle Abstufungen zwischen den Endpunkten und damit eine bessere Erfassung der Qualitäten einer Mensch-Roboter-Relation. Exemplarisch wird hierbei die Differenzierung zwischen Kooperation und Konkurrenz als Qualität der konstruierten Relation angeführt, die in der empirischen Arbeit als Spezifikum der Arbeitsplatzsituation dezidiert Anwendung fand und v. a. die virulente Angst vor Arbeitsplatzverlust als einflussreichen Faktor auf Vertrauensbildungsprozesse aufgreift. Versuche der Operationalisierung der Mensch-Roboter-Relation zu deren empirischer Untersuchung bringen die Herausforderung mit sich, die relevanten Qualitäten einer solchen Relation zu identifizieren und messtheoretisch bestimmbar zu machen. Wenngleich sich im Rahmen dieser Dissertation keine vertiefte philosophische Betrachtung der Mensch-Roboter-Relationen leisten lässt, ergeben sich aus den Ausführungen produktive Anknüpfungspunkte für weiterführende Forschungsarbeiten und mögliche Brückenschläge zwischen aktuellen philosophischen Strömungen und psychologischen Forschungen.

5.2.3 Kooperations- oder Konkurrenzorientierung als Merkmal der Relation

Ein Charakteristikum von Mensch-Technik-Relationen am Arbeitsplatz stellt das Spannungsverhältnis zwischen Kooperation und Konkurrenz dar. Die Brisanz dieser

Unterscheidung ergibt sich aus der virulenten Angst vor Arbeitsplatzverlust. Die Skizzierung einer Konkurrenzsituation zwischen Mensch und Maschine verfügt über eine gewisse historische Tradition und manifestiert sich in gängigen Narrativen von Robotern als Bedrohung der menschlichen Rolle und Position (Campolo et al., 2017; Heßler, 2019, S. 52). Ein kooperierendes Verhältnis ist gekennzeichnet von einer Haltung des Gebens und Nehmens (Reiss & Neumann, 2012). Kooperierend agierende Mitarbeitende sind davon überzeugt, einen Vorteil aus der Zusammenarbeit mit dem Cobot ziehen zu können und zeigen dafür eine gewisse Anpassungsbereitschaft. Ein Cobot erscheint als eine willkommene Unterstützung und Ergänzung der eigenen Leistungsfähigkeit. Eine dispositionelle Neigung zur Kooperation wird ferner mit einer hohen Vertrauensdisposition assoziiert. Bei einem konkurrierenden Verhältnis steht hingegen weniger die Erreichung eines gemeinsamen Ziels im Vordergrund, sondern der Wettbewerb zwischen beiden Relationspartner:innen, was häufig zu einer misstrauischen Grundhaltung führt (Fehr, 2010, S. 120f.). Solche Relationen kennzeichnen sich durch den beständigen kompetitiven Vergleich der Leistungsfähigkeit, durch ein gegenseitiges Überbietenwollen und durch die Angst des Übertroffen- und Ersetztwerdens. Die im Rahmen von Experimentalstudie 1 vorgenommene Operationalisierung des im Arbeitsplatzkontext oft thematisierten Spannungsverhältnisses zwischen Bereicherung und Bedrohung auf einem Kontinuum zwischen Kooperations- und Konkurrenzorientierung hat ein prominentes Vorbild innerhalb der *big five*-Persönlichkeitsmerkmale. In diesem Kontext wird das Merkmal der Verträglichkeit ebenfalls auf einem Kontinuum von *Kooperation* bis hin zu *Wettbewerb/Konkurrenz* gemessen (Müller & Richert, 2018; Rammstedt & Danner, 2017). Bei dieser Betrachtungsweise der Qualitäten einer Mensch-Cobot-Relation bestehen allerdings durchaus Schnittmengen zwischen den Endpunkten, die sich demnach nicht perfekt diametral gegenüberstehen. Für Relationen, die sowohl über kooperative als auch konkurrierende Anteile verfügen, existiert im Bereich des Innovationsmanagements bspw. der Begriff der „Kooppkurrenz“ (*coopetition*) (Reiss & Neumann, 2012, S. 48). Analog könnte im MRI-Kontext ein primär als Konkurrenz für den Arbeitsplatz empfundener Cobot gleichzeitig ein willkommenes Gegenüber sein, das zunächst die Arbeit erleichtert und damit auf gewisse Weise beide mentalen Modelle anspricht. Wenngleich solche Schnittmengen aus methodischer Sicht wenig wünschenswert sind, stellt ihre Vermeidung eine große Herausforderung dar, wie aktuelle Diskussionen im Zusammenhang mit dem populären Messinstrument der *Godspeed*-Skala zeigen (vgl. Kaplan, Sanders & Hancock, 2021).

Zu welchem Grad ein Roboter als Bereicherung oder Bedrohung empfunden wird, prägt die Roboterwahrnehmung und -akzeptanz maßgeblich (Meißner et al., 2020). Ein kooperierendes oder konkurrierendes Verhältnis setzt dabei nicht die Annahme voraus, dass dem Roboter ein absichtsvolles, etwa altruistisches oder egoistisches Verhalten zugeschrieben wird, sondern ergibt sich in erster Linie aus dem organisationalen Einsatzkontext. Entsprechend sind für die Wahrnehmung der Mensch-Cobot-Relation nicht primär die Fähigkeiten des Cobots maßgeblich, sondern die Einschätzung der Mitarbeitenden, was die Einführung

dieses Cobots an deren Arbeitsplatz bedeutet und welche Ziele die Unternehmensleitung damit verfolgt. Insofern ist auch an dieser Stelle zu bedenken, dass es sich beim Mensch-Cobot-Verhältnis nicht um eine dyadische Relation i. e. S. handelt, sondern um ein Vertrauensverhältnis, das bei genauerer Betrachtung ein weiter gefasstes Personengeflecht einbezieht (vgl. Kapitel 3.4.5).

Bislang liegt nur spärliche empirische Evidenz dazu vor, wie Mitarbeitende ihre Ersetzung durch Technologie bzw. die Angst davor psychisch verarbeiten und wie sich dies auf die wahrgenommene Relation zu einem Cobot auswirkt. Neben der faktischen ökonomischen Bedrohung durch den Verlust des Arbeitsplatzes steht das Selbstverständnis als leistungsfähige:r und effektive:r Mitarbeiter:in auf dem Spiel (Granulo et al., 2019). Dieses Selbstverständnis ist auch dann bedroht, wenn sich Mitarbeitende zu Zuarbeiter:innen des Roboters oder zu stumpfsinnigen Überwacher:innen der Funktionen einer Maschine degradiert fühlen (Smids et al., 2020). Die Ergebnisse aus der empirischen Arbeit zeigen, dass diese Angst unter der Belegschaft von Produktionsunternehmen virulent, salient und einflussreich ist. Daneben existieren weitere subtilere Ängste vor einer negativen Veränderung der Arbeitssituation. Die Einführung von Cobots kann in diesem Sinne einerseits gar dazu führen, dass die Mitarbeitenden weniger Gelegenheiten zu einer als sinnvoll empfundenen Beschäftigung wahrnehmen (Nyholm & Smids, 2020), andererseits aber auch die Tätigkeiten der Mitarbeitenden aufwerten und damit zu einem als sinnvoller empfundenen Arbeitsalltag beitragen (Smids et al., 2020). Verschiedene Mitarbeitende schätzen dabei den wahrgenommenen Wert einer spezifischen Tätigkeit völlig unterschiedlich ein und empfinden mitunter als unerwünscht geltende monotone Arbeiten als willkommene kognitive Entlastung. Entsprechend variantenreich ist das Spektrum an möglichen Mensch-Cobot-Relationen, in deren Rahmen der Cobot sowohl als Unterstützung bei der Arbeit oder als Bedrohung für den Arbeitsplatz gesehen werden kann. Wenngleich in der Literatur mitunter anerkannt wird, dass sich durch Cobots eine andere Dynamik der MRI entwickelt, vertreten Forschende dennoch häufig die These, dass sozial-relationale Aspekte am industriellen Arbeitsplatz erst dann relevant werden, wenn Roboter über fortschrittlichere Fähigkeiten zur sozialen Interaktion verfügen (H. Choi & Swanson, 2021). Dieser Argumentation liegt ein stark ontologischer Blickwinkel zugrunde, der erkennt, in welchem Ausmaß Cobots bereits heutzutage aus einer phänomenologisch-relationalen Perspektive etwa als Kooperationspartner:innen, als Kolleg:innen und als soziale Interaktionspartner:innen erscheinen können.

Die mit diesen Relationskonstruktionen verbundenen Erwartungshaltungen sind bisher kaum empirisch untersucht. Sie könnten ähnlich wie diejenigen Erwartungen, die sich aus der wahrgenommenen Menschenähnlichkeit ergeben, vielschichtig ausfallen, wie die Interpretation der Ergebnisse aus Experimentalstudie 1 bereits andeutet (vgl. Kapitel 4.5.4). Letztlich bilden sich die Erwartungen im Zusammenspiel aus verschiedenen Faktoren der Wahrnehmung des Cobots. So liefert bspw. eine aktuelle Studie von Lefkeli, Ozbay,

Gürhan-Canli und Eskenazi (2020) experimentelle Hinweise darauf, dass Versuchspersonen je nach Zusammenspiel aus einem vorgegebenem kooperierenden oder konkurrierenden Interaktionskontext und dem Ausgang eines gemeinsamen Spiels dem beteiligten Roboter in unterschiedlichem Maße einen Verstand zuschrieben (*mind attribution*). Das verschafft bereits einen Eindruck davon, wie komplex sich Vorhersagen über Vertrauensentwicklungen im Zeitverlauf dynamischer Interaktionssituationen gestalten. Am Umgang mit beobachteten Fehlern eines Cobots lässt sich dies exemplarisch verdeutlichen. Die Forschung zum Anthropomorphismus konnte zeigen, dass die Beobachtung von Fehlern eines Roboters bei Menschen, die diesen Roboter vergleichsweise menschenähnlich repräsentieren, zu weniger starken Vertrauensreduktionen führt (vgl. Kapitel 3.2.4.3). Der mögliche Rückschluss, dass Anthropomorphisierung die Vertrauensresilienz gegenüber Fehlern des Roboters grundsätzlich erhöht, könnte allerdings einen moderierenden Effekt der Mensch-Cobot-Relation übersehen. So wäre es denkbar, dass Fehler eines Roboters bei hoher Kooperationsorientierung zu einem Vertrauensverlust führen, weil sich die Erwartung abschwächt, dass ein solcher Roboter bei der gemeinsamen Bewältigung der Arbeitsaufgaben dienlich ist. Gleichsam könnte sich bei einer konkurrenzorientierten Relation gar ein vertrauenssteigernder Effekt durch Fehler des Roboters einstellen, da aus der beobachteten Fehleranfälligkeit eine weniger starke Bedrohung des eigenen Arbeitsplatzes resultiert. Diese hypothetischen Zusammenhänge verdeutlichen die multidimensionalen Aus- und Wechselwirkungen verschiedener Arten der Mensch-Cobot-Relation.

5.3 Anthropomorphismus

5.3.1 Impliziter Anthropomorphismus am Arbeitsplatz

Angesichts der erheblichen Aufmerksamkeit, die das Thema der Vermenschlichung in der Forschung erfährt, überraschte zunächst die vergleichsweise geringe Relevanz, die die Produktionsmitarbeitenden der Thematik in den Vorstudien beimaßen. Erst bei detaillierter Betrachtung der Aussagen ließen sich Hinweise auf mögliche, implizit ablaufende Vermenschlichungsprozesse finden (vgl. Kapitel 4.3.3). Diese qualitativen Eindrücke bestätigten sich quantitativ in der ersten Experimentalstudie. Mittels expliziter und in stärkerem Ausmaß mittels indirekter Messmethoden ließen sich vermenschlichende Zuschreibungen – vermutlich begünstigt durch entsprechendes Framing – identifizieren (vgl. Kapitel 4.5.3.1). Dass die Effekte bei der indirekten Messung stärker zum Vorschein traten, lässt sich mit der Theorie des dualen Anthropomorphismus erklären (vgl. Kapitel 3.2.3.2). Die grundsätzliche Annahme dabei lautet, dass die indirekte Messung diejenigen vermenschlichenden Zuschreibungen offenlegt, die im Zuge eines automatisch und unbewusst ablaufenden Typ-1-Verarbeitungsprozesses entstehen. Die gezielte Befragung nach der Menschenähnlichkeit eines Roboters initiiert hingegen die aktive und bewusste kognitive

Beschäftigung mit der Fragestellung im Rahmen eines elaborierten Typ-2-Verarbeitungsprozesses. Zunächst entsteht dadurch eine Dissonanz zwischen den affektiven Wahrnehmungen und den tradierten und sozial akzeptierten Überzeugungen (Coeckelbergh, 2018, S. 154). Diese löst sich dadurch auf, dass der Typ 2-Verarbeitungsprozess die automatischen vermenschlichenden Zuschreibungen korrigiert. In expliziten Antworten auf die Frage nach der Menschenähnlichkeit lassen sich Vermenschlichungen entsprechend nur noch in geringerem Maße nachweisen. Diese liefern damit eher Einsichten in das Vorhandensein von starkem Anthropomorphismus und beantworten die Frage, ob die jeweilige Person tatsächlich davon überzeugt ist, dass das technische Artefakt aus einer deskriptiven Perspektive einem Menschen gleicht bzw. aus einer normativen Perspektive eine vergleichbare Behandlung verdient⁷. Der stärker nachweisbare implizite Anthropomorphismus stellt hingegen ein Indiz dafür dar, dass der Roboter intuitiv und affektiv menschenähnlich *erscheint*. Gerade im funktional geprägten Arbeitsplatzkontext könnte das explizite Vertreten starker Positionen des Anthropomorphismus vor dem Hintergrund der Leistungsorientierung und im Beisein anderer Unternehmensmitarbeitender als unüblich und wenig akzeptiert gelten. Desto mehr legen die empirischen Befunde nahe, fehlende Nachweise von Anthropomorphismus in direkten Messmethoden nicht als Indiz dafür zu interpretieren, dass überhaupt keine Vermenschlichung stattfindet. Vielmehr ist anzuraten, ergänzend indirekte Messmethoden wie bspw. die im Rahmen der Experimentalstudien verwendete linguistische Analyse von Äußerungen einzusetzen, um impliziten Anthropomorphismus offenzulegen.

5.3.2 Kontextabhängige Auswirkungen von Anthropomorphisierung

Gerade im Kontext der sozialen Robotik gilt eine starke Vermenschlichung eines Roboters als vertrauensstiftender Faktor (Zawieska et al., 2012). Allerdings offenbart eine genauere Betrachtung empirischer Studien konfligierende Forschungserkenntnisse (vgl. Kapitel 3.2.4.3). So verweisen Roesler und Onnasch (2020b, S. 172f.) auf kontroverse empirische Auswirkungen von Anthropomorphisierung in realistischen industriellen MRI-Kontexten. Ferner fordern sie, den Wunsch nach allgemeingültigen Gestaltungsempfehlungen angesichts der Abhängigkeit dieser Effekte vom Einsatzgebiet und von der Aufgabe des Roboters aufzugeben. Unterschiedliche Arten der Operationalisierung sowie zu wenig berücksichtigte Interaktions- und Rollenkontexte verringern die Vergleichbarkeit einzelner experimenteller Ergebnisse (Onnasch & Roesler, 2021).

⁷ Zum Unterschied zwischen deskriptiv-sacherläuternder vermenschlichender Rede über Roboter in Form von *so wie*-Metaphern im Vergleich zu normativer Rede in Form von *als ob*-Metaphern siehe Nerurkar, Wadephul und Wiegerling (2019) sowie Wadephul (2020).

Die bekannte Theorie des unheimlichen Tals (vgl. Kapitel 3.2.4.4) umschreibt bereits, wie sich zunächst positive Effekte eines stärkeren anthropomorphen Designs ab einem gewissen Grad umkehren. Kontextabhängig können sich weitere unerwartete und nicht-intendierte Anthropomorphismuseffekte einstellen. So weisen Yogeewaran et al. (2016, S. 40) bspw. darauf hin, dass die menschenähnliche Gestaltung eines Roboters „nach hinten losgehen“ (*backfire*) kann, wenn bei den betroffenen Personen der Eindruck entsteht, dass dieser Roboter sie in ihren Leistungen übertreffen könnte. Auch in einer jüngeren Meta-Analyse mit einem stärkeren Fokus auf aufgabenorientierten Interaktionen zeigten sich überraschenderweise keine signifikanten Effekte von Vermenschlichung auf Vertrauen (Ötting et al., 2020). Das führen die Autor:innen ebenfalls auf eine starke Aufgaben- und Kontextabhängigkeit der Effekte von Vermenschlichung zurück. Ferner vermuten sie, dass Anthropomorphisierung bei für den (industriellen) Arbeitsplatz typischen leistungsorientierten Interaktionen eine weniger gewichtige Rolle als bei sozialen Interaktionen spielt (Ötting et al., 2020, S. 5/12).

Die zuvor eingeführte Differenzierung in leistungs- und relationsorientiertes Vertrauen (vgl. Kapitel 5.1.3) bietet einen geeigneten Ansatzpunkt, um diese These näher zu beleuchten und theoretisch zu untermauern. Wie bereits geschildert, stehen am industriellen Arbeitsplatz leistungsorientierte Vertrauensaspekte im Vordergrund (Malle & Ullman, 2021), ohne dass relationsorientierte Komponenten vollends ihre Bedeutung verlieren (vgl. die Zielsetzungen auf Basis der Drei-Faktoren-Theorie in Kapitel 4.5.4.2). Vermenschlichungseffekte sind in diesem Kontext in erster Linie insofern relevant, als sie Leistungserwartungen implizieren. Je nach Aufgabenkontext könnte ein vermenschlichter oder ein mechanisch repräsentierter Roboter als geeigneter betrachtet werden, weil jeweils entweder spezifisch menschliche Fähigkeiten wie Flexibilität und Kreativität oder maschinelle Fähigkeiten wie hohe Präzision und Wiederholgenauigkeit erforderlich sind. Neben der grundsätzlichen Eignung für die jeweilige Aufgabe führt eine starke Vermenschlichung zu einer geringeren Aktivierung des PAS (vgl. Kapitel 3.2.4.3). Das hat eine niedrigere Zuverlässigkeitserwartung und eine höhere Fehlertoleranz zur Folge. Aus der niedrigen Zuverlässigkeitserwartung ergibt sich wiederum ein geringeres initiales Vertrauen in Hinblick auf das Ziel, mithilfe des Roboters gute Arbeitsergebnisse erzielen zu können. Ein hohes Maß an Vermenschlichung kann sogar dazu führen, dass der funktionale Mehrwert des Roboters völlig in den Hintergrund tritt (Onnasch & Roesler, 2019; Roesler & Onnasch, 2020a). Die wahrgenommene Menschenähnlichkeit wirkt sich damit über die zugeschriebenen Fähigkeiten auf die Vertrauenswürdigkeit des Cobots aus.

Überdies ist allerdings auch eine Auswirkung auf das relationsorientierte Vertrauen zu vermuten, da ein stark vermenschlichter Roboter das Vertrauen darin bestärkt, als soziale:r Interaktionspartner:in aufzutreten (vgl. die in Kapitel 3.2.3.3 beschriebene Sozialitätsmotivation). Abbildung 5.5 visualisiert diese hypothetischen Zusammenhänge zwischen dem Interaktionskontext, dem zweidimensionalen Vertrauen, den Kriterien für

im Kontext der industriellen MRI. Darüber hinaus stellt er ein starkes Indiz dafür dar, dass Menschen die (drohende) Ersetzung durch einen anderen Menschen oder durch eine Technologie psychisch unterschiedlich verarbeiten und ihren Selbstwert dadurch unterschiedlich stark gefährdet sehen (Granulo et al., 2019). Das dargestellte Modell bietet einen ersten Ansatzpunkt, um die kontroversen empirischen Befunde zu den Auswirkungen von Anthropomorphismus auf Vertrauen unter Berücksichtigung multidimensionaler Zusammenhänge, des Interaktionskontexts und der Mensch-Cobot-Relation besser zu erklären. Gleichwohl bedarf es einer vertieften Ausarbeitung und empirischen Prüfung in zukünftiger Forschung.

5.4 Sprachliches Framing

5.4.1 Prinzipielle Wirkkraft von striktem Framing

Ein Ziel der empirischen Untersuchungen bestand in der Klärung der Frage, ob sich subtile Variationen der vermittelten sprachlichen Beschreibungen eines Cobots auf die wahrgenommene Menschenähnlichkeit sowie auf die Mensch-Cobot-Relation und damit mittelbar auf das Mensch-Cobot-Vertrauen auswirken. Die bisher vorliegende Evidenz ließ solche Effekte vermuten, war allerdings noch wenig verdichtet und bezog sich ausschließlich auf den Aspekt der Menschenähnlichkeit (vgl. Kapitel 3.3.3). Die Ergebnisse aus Experimentaltstudie 1 verdeutlichen signifikante Veränderungen im Grad der Vermenschlichung sowie der Kooperationsorientierung der Mensch-Cobot-Relation in Abhängigkeit vom jeweiligen Framing. Die Effektgrößen lagen statistisch gesehen in einem kleinen bis mittelstarken Bereich bei der wahrgenommenen Menschenähnlichkeit bzw. im mittelstarken bis starken Bereich bei der Kooperationsorientierung der Relation. Letztere lässt sich aufgrund der eigens entwickelten Skala nicht mit anderen Studien vergleichen. Die Effektstärke in Bezug auf die wahrgenommene Menschenähnlichkeit ist vor dem Hintergrund einer vermuteten geringen Sensitivität der verwendeten *Godspeed*-Skala bemerkenswert (vgl. Kapitel 4.5.4.1). In früheren Studien konnten mittels der *Godspeed*-Skala selbst beim Vergleich eines tatsächlich anthropomorph gestalteten mit einem maschinell gestalteten Roboter mitunter keine Veränderungen in der wahrgenommenen Menschenähnlichkeit ermittelt werden (siehe z. B. Roesler & Onnasch, 2020a). Wenngleich die *Godspeed*-Skala eine verbreitete Methode zur Operationalisierung darstellt, ist eine Kontextualisierung der gemessenen Wirkkraft des Framings vor dem Hintergrund ähnlicher Studien kaum möglich, da die unterschiedlichen Definitionen von Framing sowie verschiedene Versuchsetzungen die Vergleichbarkeit einschränken (Stocké, 2002, S. 23).

Die Ergebnisse der vorliegenden Dissertation weisen darauf hin, dass nicht nur Unterschiede in den vermittelten Informationen, sondern auch in deren Ausformulierung die

mentalen Modelle beeinflussen. Das ist besonders bemerkenswert, da erstens die vermittelten Informationen insgesamt wenig umfangreich waren, zweitens die Variation moderat ausfiel, und drittens gerade die expliziten Erhebungsmethoden möglicherweise nur wenig sensitiv auf subtile Veränderungen in den impliziten mentalen Modellen reagieren. Die Variation bestand in der Verwendung von Formulierungen (lexikalische Ebene), die entweder dem menschlichen oder dem maschinellen Begriffsraum zuzuordnen waren, in der Nutzung von aktiver oder passiver Sprache (syntaktische Ebene) sowie in der sprachlichen Akzentuierung bestimmter Informationen und Bedeutungen im Sinne sog. Hervorhebungsframes (semantische Ebene). Insofern handelt es sich um Framing im strikten Sinne, das sich die Interpretationsbedürftigkeit von Sprache zunutze macht und gezielt bei der Art der Informationsvermittlung ansetzt (Stocké, 2002, S. 41). Entsprechend sind bei der Interpretation der Ergebnisse andere Maßstäbe anzulegen als bei thematisch ähnlichen Studien, die allerdings als experimentelle Manipulation gänzlich unterschiedliche Informationen über Roboter vermittelten, wie bspw. abweichende Zuverlässigkeitswerte in Form quantitativer Prozentangaben (vgl. Kapitel 3.3.3). Zu beachten ist hierbei, dass die in den Texten verwendeten Merkmale konsistent entweder dem Begriffsraum des Menschlichen oder des Maschinellen entstammten. Damit sollten sie einen eindeutigen Interpretationsrahmen repräsentieren, der konsistent die Aktivierung eines bestimmten mentalen Modells fördert und somit eine Methode zur Disambiguierung der vorgefundenen Interaktionssituation darstellt.

Bei der konkreten sprachlichen Ausgestaltung des Framings bestehen zahlreiche Freiheitsgrade. So wurde der Cobot beim menschenähnlichen Framing mit dem männlichen und in Deutschland gebräuchlichen Namen *Paul* bezeichnet. Insofern findet unvermeidlich gleichzeitig ein Framing des Geschlechts und der Nationalität des Cobots statt. Insbesondere die Auswirkungen von geschlechtsspezifischen Stereotypen in Bezug auf die Wahrnehmung eines Roboters wurden bereits in einigen Studien untersucht (Appel et al., 2020; Eyssel & Hegel, 2012; Law, Chita-Tegmark & Scheutz, 2021; Siegel et al., 2009; Tay, Jung & Park, 2014). Die Berücksichtigung dieser Ergebnisse in weiterführenden Studien würde eine genauere Ausdifferenzierung in Hinblick darauf ermöglichen, welche sprachlichen Elemente innerhalb des Framings welche mentalen Modelle und damit einhergehenden Erwartungen aktivieren. Es ist davon auszugehen, dass sich eine abweichende Wirkung des Framings einstellen würde, wenn die Cobot-Beschreibung Formulierungen aus verschiedenen Begriffsräumen enthielte, die ihrerseits eine weitere Ebene der (semantischen) Ambiguität hinzufügen.⁸ Bei der Übertragung und Anwendung von Termini aus einem bestehenden Sprachraum auf ambige Technologien können in diesem Zusammenhang „Fallstricke“ (*pitfalls of ambiguous terminology*) (Remmers, 2019, S. 59ff.) vorkommen, die letztlich zur Manifestation von Fehlkonzeptionen führen. Ein Beispiel hierfür stellt die

⁸ Zu verschiedenen Formen sprachlicher Ambiguität siehe z. B. M. Bauer et al. (2010, S. 47ff.).

Rede von autonomen Robotern dar. Während der Begriff der Autonomie in Bezug auf Menschen über eine lange Forschungstradition, eine tiefere Bedeutung und eine normative Komponente verfügt, ist der Begriff in Bezug auf Roboter wesentlich weniger aufgeladen und bezieht sich allein auf das eigenständige Handeln eines Roboters ohne eine:n Operator:in (Remmers, 2019). Diese unterschiedlichen Konnotationen je nach Anwendungsgebiet können leicht zu missverständlichen Schlüssen über die Fähigkeiten von Robotern führen.

Für die Existenz praxisrelevanter Auswirkungen des Framings spricht ein messbarer signifikanter Effekt auf das Vertrauen. Trotz des bemerkenswerten Einflusses auf das Vertrauen erwies sich gleichwohl die Grundeinstellung gegenüber Robotern unabhängig von der Framing-Bedingung als aussagekräftiger Prädiktor für das empfundene Vertrauen. Mit steigender Interaktionsfähigkeit von Cobots kommt dem Framing und dem Erwartungsmanagement im Vorfeld einer Interaktion perspektivisch eine höhere Bedeutung zu, weil Cobots damit in noch höherem Maße ambige Stimuli darstellen (Washburn, Adeleye, An & Riek, 2020, S. 5). In Ergänzung zum Vorgehen in den Experimentalstudien ist ferner denkbar, dass Framing während der MRI zur Anwendung kommt, bspw. um Fehler des Roboters unterschiedlich zu erklären und damit radikalen Vertrauensverlusten vorzubeugen. Hierzu sind weitere Forschungen zur Übermittlung von Transparenzinformationen bei unerwarteten Fehlern des Roboters relevant (vgl. u. a. Miller, 2021).

5.4.2 Wahrgenommene Ambiguität als Erklärungsansatz

Nachdem im vorigen Kapitel die prinzipielle Wirkkraft von Framing erläutert wurde, stellt sich daran anknüpfend die Frage, wie sich erklären lässt, dass sich geringe textuelle Variationen im beschriebenen Ausmaß auf die Wahrnehmung auswirkten. Erneut scheint die (wahrgenommene) Ambiguität des Cobots einen entscheidenden Schlüssel zum Verständnis darzustellen. Im Folgenden werden u. a. Aspekte des MdFS herangezogen, um die Wirkung von Framing in diesem speziellen Kontext theoretisch zu erklären (siehe hierzu auch Kapitel 3.2.3.4). Framing wird hierbei als situative Aktivierung je verschiedener mentaler Modelle verstanden, die in der Folge die Basis für Schlüsse über unbekannte Merkmale und künftige Aktionen des Cobots bilden. Abbildung 5.6 visualisiert schematisch die Abfolge der entsprechenden kognitiven Prozesse.

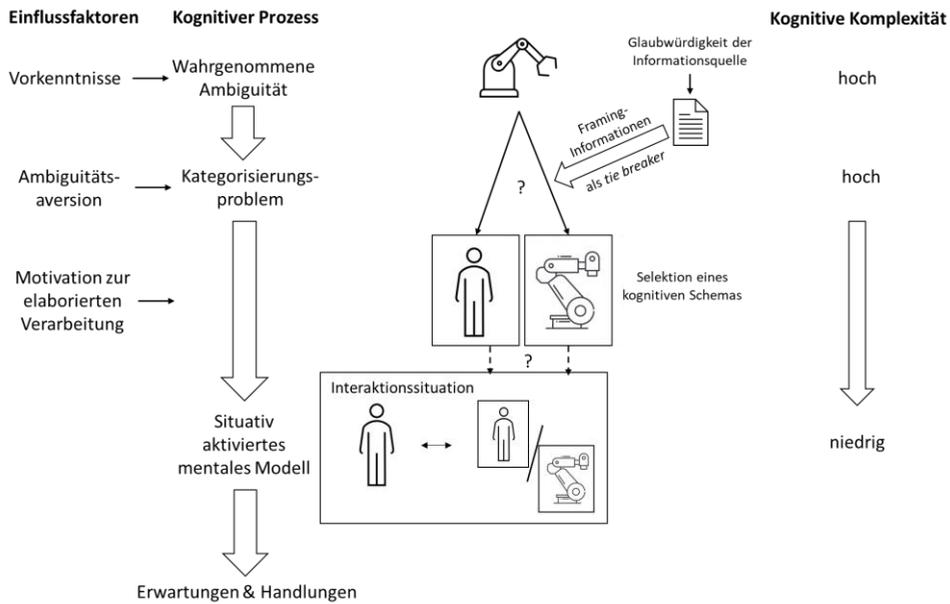


Abbildung 5.6: Schematische Darstellung des kognitiven Prozesses bei einer tatsächlichen oder bevorstehenden Konfrontation mit einem Cobot und des Einflusses von sprachlichem Framing als *tie breaker* zur Auflösung von Ambiguität.

Die Wahrnehmung von Ambiguität begünstigt die Wirkkraft von Framing, da sie eine Versicherung zur Folge hat (Stocké, 2002, S. 37f.; Ziegler, 2010), bspw. in Hinblick auf das tatsächliche Wesen, die Erscheinung, die Fähigkeiten, die mentalen Zustände und die künftigen Aktionen eines Cobots. Verschiedene Zukunftsentwürfe müssen parallel verarbeitet und mental simuliert werden, was eine hohe kognitive Belastung verursacht. In diesem Fall erfüllt die heuristische Anwendung vorgefertigter kognitiver Schemata zur Bildung eines situativen mentalen Modells den Anspruch der kognitiven Ökonomie. Wenn der Cobot aufgrund mangelnder oder widersprüchlicher Vorerfahrungen und Vorkenntnisse weder im Gesamten als prototypisch für ein kognitives Schema angesehen wird noch durch selektive Wahrnehmung einzelne eindeutige Merkmale ins Zentrum der Aufmerksamkeit rücken, gelingt es zunächst nicht, ein salientes kognitives Schema zu identifizieren. In der Sprache des MdFS würde das bedeuten, dass ein Selektionsproblem vorliegt (vgl. Kapitel 3.2.3.4). Ansätze aus der Psychologie rekurrieren in diesem Zusammenhang auf den Begriff des Kategorisierungsproblems (Burleigh & Schoenherr, 2015). Dieses Problem kann sich sowohl auf die Unterscheidungsdimension zwischen Menschen und Maschinen als auch auf die relationsorientierte Dimension zwischen Kooperationspartner:innen und Konkurrent:innen beziehen.

Im Rahmen des vorgestellten Erklärungsansatzes kommt dem sprachlichen Framing die Rolle des sprichwörtlichen *Züngleins an der Waage* zu, das den entscheidenden Ausschlag dafür gibt, in welche Richtung die Mehrdeutigkeit aufgelöst wird und in welche Richtung die Wahrnehmung ausschlägt. Stocké (2002, S. 56f.) spricht in diesem Zusammenhang von einer sog. *tie breaker*-Hypothese, die einigen Framing-Theorien zugrunde liegt und impliziert, dass Framing in Situationen relevant ist, in denen Entscheider:innen den Wahlmöglichkeiten indifferent gegenüberstehen. Tendenziell ist davon auszugehen, dass mit einem größeren Ausmaß an Verunsicherung und damit verbundener kognitiver Belastung die Bereitschaft wächst, sprachliche Frames weitestgehend unreflektiert als Hinweis für die Aktivierung eines menschenähnlichen oder maschinellen bzw. eines kooperierenden oder konkurrierenden mentalen Modells heranzuziehen. Die Stärke der Framing-Effekte ist daher ein Indiz für die wahrgenommene Ambiguität der Stimuli und der Unsicherheit über die Kategorisierung des betreffenden Roboters. Auch der Effekt des unheimlichen Tals und die damit verbundenen aversiven Reaktionen könnten darauf zurückzuführen sein, dass bei einem relativ hohen Maß an Menschenähnlichkeit keine eindeutige Einschätzung darüber möglich ist, ob es sich tatsächlich um einen Menschen oder eine Maschine handelt (Samuel, 2019; Smith & Wiese, 2016).

Die Unsicherheit ob der mentalen Einordnung eines Cobots spiegelte sich bereits qualitativ in den Befragungen von Produktionsmitarbeitenden in den Vorstudien sehr deutlich wider und resultierte auch dort in einer hohen Aufmerksamkeit und Sensitivität für die vermittelten Informationen (vgl. Kapitel 4.4). Dass sich anthropomorphe und kooperationsorientierte mentale Modelle durch die geringfügigen Variationen in den Framing-Texten auch im arbeitsplatzbezogenen Kontext der industriellen MRI aktivieren ließen und sich auf das initiale Vertrauen auswirkten, steht dem teils vermittelten Eindruck entgegen, dass Vermenschlichung und sozial-relationale Aspekte in diesem funktionalen Kontext grundsätzlich (noch) keine Rolle spielen (H. Choi & Swanson, 2021, S. 132). Aufgabenbasierte Tätigkeitskontexte stellen demnach keine übergeordnete Rahmung dar, die anthropomorphe Repräsentationen des Cobots gänzlich verhindert.

5.4.3 Abhängigkeit von personenbezogenen Faktoren

Der in Abbildung 5.6 skizzierte Erklärungsansatz benennt bereits drei personenbezogene Einflussfaktoren auf die Wirkkraft des Framings, nämlich die vorhandenen Vorkenntnisse, die individuelle Ambiguitätsaversion und die Motivation zur elaborierten Verarbeitung. Anhand dieser Einflussfaktoren lässt sich die beobachtete Diskrepanz in der Wirkung des Framings in beiden Experimentalstudien theoretisch erklären.

Zunächst kann der Besitz entsprechenden Vorwissens in Bezug auf Roboter dazu führen, dass eine vorgefundene Situation nicht als ambig wahrgenommen wird, wenn die

entsprechenden Personen Roboter als prototypisch für ein bestimmtes kognitives Schema empfinden und kategorisch menschenähnlich oder maschinell repräsentieren. In der Sprache der MdFS ausgedrückt, würde „in der betreffenden Situation das [zum jeweiligen mentalen Modell] zugehörige signifikante Symbol“ (Esser, 2005, S. 7) vorkommen, das eine heuristische Aktivierung des mentalen Modells veranlasst. Demnach wäre es plausibel, dass die Studierenden aus Experimentalstudie 2 im Rahmen ihres Studiums bereits ein sehr klar technisiertes Bild von Robotern vermittelt bekommen haben und daher dieses maschinelle mentale Model anwendeten. In diesem Fall hätten sie bei der Konfrontation mit dem Cobot allenfalls in geringem Maße Ambiguität empfunden. Ein hohes Maß an wahrgenommener Mehrdeutigkeit ist allerdings konstitutiv dafür, dass vermittelte Framing-Informationen zur Entscheidung zwischen verschiedenen mentalen Modellen herangezogen werden. Insofern lässt sich die Abhängigkeit der Stärke von Framing-Effekten vom Vorwissen der betreffenden Personen erklären. Es sei allerdings darauf hingewiesen, dass der Begriff des Vorwissens insofern missverständlich ist, als es sich hierbei nicht zwangsläufig um tatsächliches Faktenwissen in einem objektiven Sinne handelt, sondern in letzter Konsequenz die Überzeugung der jeweiligen Person entscheidend ist, eindeutiges Wissen über das Wesen eines Roboters zu besitzen.

Die tendenziell aversiven Reaktionen auf die wahrgenommene Ambiguität wurden bereits in Kapitel 5.1.2 thematisiert. Diese spielen bereits in den klassischen Experimenten zur Entscheidungstheorie eine wichtige Rolle (Ellsberg, 1961; Tversky & Kahneman, 1981). Bisher unternahmen allerdings nur wenige Studien den Versuch, die Stärke der Ambiguitätsaversion – oder umgekehrt formuliert: die individuelle Ambiguitätstoleranz – als Einflussfaktor auf Framing-Effekte zu untersuchen (Osmont, Cassotti, Agogué, Houdé & Moutier, 2015). Das Konzept der Ambiguitätstoleranz bezieht sich auf die individuell unterschiedliche Reaktion auf Ambiguitätswahrnehmungen sowie auf deren Bewertung. Ambiguitätsintolerante Menschen reagieren affektiv mit Stress und Unsicherheit auf ambige Situationen und versuchen daher, diese zu meiden oder die Inkongruenz dargebotener Stimuli zu verdrängen. Solche Personen haben i. d. R. ein starkes Sicherheitsbedürfnis und tendieren in hohem Maße zu einer Dichotomisierung in vorgefertigte Denkkategorien. Ambiguitätstolerante Personen hingegen empfinden solche Situationen als angenehm herausfordernd und haben damit nur eine geringe Motivation zur Disambiguierung (Ziegler, 2010, S. 144ff.). Insofern die kognitive Komplexität der Mehrdeutigkeit für ambiguitätstolerante Personen keine Belastung darstellt, besteht für sie nur eine geringe Veranlassung, zusätzliche vage Informationen aus dem Framing zur Auflösung der Ambiguität heranzuziehen. Hierbei spielt überdies die Glaubwürdigkeit (*credibility*) eine entscheidende Rolle, die der jeweilige Rezipient der Informationsquelle zuschreibt. Wenn die Informationen als wenig relevant oder wenig glaubwürdig angesehen werden, ist davon auszugehen, dass insbesondere ambiguitätstolerante Menschen diese ignorieren und die Mehrdeutigkeit damit bestehen lassen. Da Glaubwürdigkeit ein wichtiger Bestandteil der Vertrauenswürdigkeit darstellt (Madhavan & Wiegmann, 2007, S. 285), lässt sich schlussfolgern, dass ein

etwaiges Misstrauen in die Quelle der Framing-Informationen zur Manifestation der Ambiguität führt und damit die Vertrauensbildung in Hinblick auf das jeweilige technische System behindert.

Ferner werden eine situativ hohe individuelle Motivation, sich intensiv und elaboriert mit einer Thematik zu beschäftigen und die situativ dafür verfügbaren kognitiven Fähigkeiten als Faktoren diskutiert, die die Wirksamkeit von Framing verringern (Stocké, 2002, S. 157). Die Motivation kann sich dabei intrinsisch aus einem personenindividuell unterschiedlich stark ausgeprägten allgemeinen Kognitionsbedürfnis (*need for cognition*) ergeben (Cacioppo, Petty, Feinstein & Jarvis, 1996) oder aus der Überzeugung, dass die zu treffende Entscheidung wichtig ist (Stocké, 2002, S. 84f.). Im Rahmen der MdFS, die in einem *dual process*-Ansatz von zwei unterschiedlichen Verarbeitungsmodi ausgeht, wird in diesem Zusammenhang der Begriff der Reflexionsmotivation verwendet (vgl. Kapitel 3.2.3.4). Im Regelfall erfolgt die Aktivierung passender kognitiver Schemata automatisiert-spontan. Für diesen schnell ablaufenden Prozess können selektiv verfügbare Framing-Informationen entscheidend sein. Wird die Entscheidung allerdings als wichtig erlebt oder werden etwa von Dritten Begründungen hierfür eingefordert, kann dies zur Aktivierung eines stark kognitiv geprägten reflektierenden Verarbeitungsmodus führen. Die Aktivierung eines entsprechenden mentalen Modells geschieht dann auf Basis tiefergehender elaborierter Denkabwägungen und zeigt sich robuster gegen situative Einflüsse wie subtile Framing-Informationen (Esser, 2005).

Die unterschiedliche Wirkkraft des Framings auf die beiden Stichproben in den Experimentalstudien lässt sich zusammenfassend vor dem Hintergrund dieser Einflussfaktoren verstehen. Der Vergleich der Stichproben offenbart bereits ein höheres (wahrgenommenes) Vorwissen der Studierenden (vgl. Kapitel 4.7.1). Ferner verfügen Studierende typischerweise „über ein überdurchschnittliches Ausmaß an [Reflexionsm]otivation und kognitiven Fähigkeiten“ (Stocké, 2002, S. 257), was sie robuster gegenüber Framing-Einflüssen macht. Auch in diesem Zusammenhang lässt sich die ökologische Validität von Framing-Studien mit studentischen Teilnehmenden im Kontext der industriellen MRI problematisieren. Neben diesen personenspezifischen Einflussfaktoren ist in Hinblick auf die Experimentalstudien zu bedenken, dass Framing auf kognitiver Ebene wirkt, wohingegen in Experimentalstudie 2 durch die tatsächliche Konfrontation mit dem Cobot affektive Vertrauenskomponenten stärker angesprochen wurden.

5.4.4 Persistenz von Framing-Effekten

Die Erklärung von Framing-Effekten über die situationsbezogene Aktivierung von mentalen Modellen wirft die Frage nach der Persistenz dieser Effekte auf. Diese Frage ist zentral für die praktische Relevanz von Framing-Effekten (Ahn, Fox, Dale & Avant, 2015; Baden

& Lecheler, 2012), ließ sich allerdings im Rahmen der vorliegenden Dissertation aufgrund der unwirksamen experimentellen Manipulation in Experimentalstudie 2 empirisch nicht untersuchen. In bestehender MRI-Literatur wurden langfristige Effekte bis dato wenig untersucht, wengleich sich in neueren Forschungsarbeiten die Forderungen danach mehren (Nam et al., 2021, S. 486). Eine Ausnahme stellt die Studie von Zlotowski und Sumioka, et al. (2015) dar, die die zeitliche Persistenz von Effekten des unheimlichen Tals untersucht. Der Großteil der Studien zur Existenz von Framing-Effekten beschäftigt sich allerdings nur mit deren initialem Auftreten, ohne zu überprüfen, ob die Effekte kurze Laborsituationen überdauern (Lecheler & de Vreese, 2011, 2016).

Aufgrund dieses Mangels an verfügbarer Literatur im MRI-Kontext wird zur Abschätzung der Persistenz von Framing-Effekten an dieser Stelle auf kommunikationswissenschaftliche Theorieansätze zurückgegriffen. Grundsätzlich ist davon auszugehen, dass sich Framing-Effekte im Zeitverlauf allmählich zurückbilden, zumindest sofern es sich beim Framing um eine einmalige Intervention handelt und keine wiederholte Konfrontation mit diesem Frame stattfindet (Lecheler & de Vreese, 2016). Gleichwohl liegen empirische Studien vor, in denen sich zwar langsam zurückbildende, aber dennoch nach mehreren Wochen signifikant messbare Effekte zeigen (Lecheler & de Vreese, 2011). In einer entsprechenden Untersuchung zum Framing der politischen Orientierung führen Dharshing, Hille und Wüstenhagen (2017) die Persistenz eines Frames auf die Neigung zurück, an der bestehenden Meinung festzuhalten. Gerade im Kontext solcher Studien, die sich auf die mediale Kommunikation und die Rezeption von Nachrichtenmeldungen beziehen, kommt der Wirkung zwischenzeitlich rezipierter konkurrierender Frames eine hohe Bedeutung zu. Hierbei wird davon ausgegangen, dass die wiederholte Präsentation konsistenter Frames (repetitives Framing) zu einer Verstärkung von Framing-Effekten führt, wohingegen inhaltlich unvereinbare oder gegensätzliche Frames (*counter-frames*) aufgrund eines *recency*-Effekts die Wirkung früherer Frames überlagern können (Lecheler & de Vreese, 2016).

Zur theoretischen Erklärung der unterschiedlichen experimentellen Erkenntnisse entwickelten Baden und Lecheler (2012) ein wissensbasiertes Modell zur Persistenz von Framing-Effekten. Darin unterscheiden sie in Bezugnahme auf bestehende Ansätze verschiedene kognitive Wirkmechanismen von Framing, die sich mit dem in Kapitel 5.4.2 entwickelten Modell in Verbindung bringen lassen. Bisherige Erklärungsmuster basieren erstens auf einer leichteren Zugänglichkeit von durch Framing voraktiviertem Wissen ähnlich wie bei Priming-Prozessen (*accessibility*), zweitens auf der Anwendbarkeit verfügbaren Wissens auf die vorliegende Situation (*applicability / belief importance change*) und drittens auf einer Veränderung der Überzeugungen der jeweiligen Person (*belief content change*), die durch neue Informationen innerhalb des Framings vermittelt werden. Während sich die kognitive Zugänglichkeit schnell wieder zurückbildet, können Framing-Interventionen einen länger anhaltenden Effekt hervorrufen, wenn sie auf einem

ausgeweiteten Anwendungsbereich von Wissen oder auf einer Veränderung von Überzeugungen basieren. Dies führen die Autor:innen darauf zurück, dass mit diesen Vorgängen neue Informationen im Langzeitgedächtnis hinterlegt werden, also eine Art von Lernen stattfindet. Das Lernen kann sich darin ausdrücken, dass der Anwendungsbereich von bestehenden mentalen Modellen ausgedehnt oder sogar die Wissensstruktur partiell modifiziert wird, die innerhalb des mentalen Modells repräsentiert ist. Daraus ergibt sich ferner die Implikation, dass Framing-Effekte bei Personen mit einem gefestigten Vorwissen oder einer starken Ignoranz gegenüber dem im Framing vermittelten Wissen tendenziell weniger lange überdauern (Baden & Lecheler, 2012).

Zur Anwendung auf MRI-Situationen empfiehlt sich daher die Unterscheidung in zwei Arten von Persistenz. Von situationsbezogener Persistenz lässt sich sprechen, wenn die Wirkung von Framing-Effekten während der Dauer einer konkreten MRI-Situation bestehen bleibt. Wenn sich diese Effekte überdies auch bei weiter in der Zukunft liegenden verwandten MRI-Situationen bemerkbar machen, wäre von situationsübergreifender Persistenz zu sprechen. Situationsbezogen sind sowohl Framing-Effekte auf Basis der leichteren Zugänglichkeit, der Anwendbarkeit als auch der Überzeugungsveränderung relevant. Zu bedenken ist dabei, dass direkte Beobachtungen eines Roboters durch deren hohe Salienz zuvor vermittelte Informationen überlagern können, die zunächst das entsprechende Framing der Situation hervorgerufen hatten. Dies deckt sich mit der Beobachtung, dass der Grad der Vermenschlichung zu Beginn einer Interaktion sehr stark vom Erscheinungsbild des Roboters abhängt, sich aber rasch durch Eindrücke aus der Beobachtung verändern kann (Abubshait & Wiese, 2017). Gleichwohl ist davon auszugehen, dass Beobachtende aus Gründen der kognitiven Ökonomie zunächst versuchen, ihre neuen Eindrücke zu assimilieren, also unter Beibehaltung des gegenwärtigen mentalen Modells zu verarbeiten. Insofern müssen widersprüchliche Reize zuerst eine gewisse Salienz erreichen, um ein verändertes Framing auszulösen.

Damit Framing-Effekte situationsübergreifend wirken, ist es notwendig, dass diese auf einer Art des Lernens, also auf einer Veränderung in der Anwendbarkeit bestehenden Wissens oder auf dessen Modifikation durch das Framing beruhen. Eine situationsbezogene Kategorisierung eines Roboters als maschinell könnte bspw. zu einer Ausweitung des prototypischen Anwendungsbereiches des entsprechenden maschinellen mentalen Modells führen. Ein ähnlicher Roboter in einer späteren Interaktionssituation würde infolge des früheren Framings erneut in gleicher Weise mental repräsentiert. Das wäre auch der Fall, wenn sich durch die Informationen aus dem Framing bspw. gar die bewusste Überzeugung ausgebildet hätte, dass es sich bei sämtlichen Robotern eindeutig um rein maschinelle Entitäten handelt.

6 Implikationen, Reflexion und Ausblick

Im Folgenden werden zunächst in Kapitel 6.1 sowohl unternehmerische als auch gesellschaftliche Implikationen aufgezeigt, die sich aus den Forschungsergebnissen ergeben. Kapitel 6.2 analysiert daraufhin die Übertragbarkeit der Ergebnisse auf andere Forschungsgegenstände. In Kapitel 6.3 findet eine abschließende methodische Reflexion statt, die die Erfahrungen mit dem gewählten methodischen Ansatz und die damit verbundenen Einschränkungen darstellt. In Kapitel 6.4 folgt abschließend ein Ausblick auf mögliche anknüpfende Forschungsarbeiten.

6.1 Implikationen

6.1.1 Unternehmerische Perspektive

Hinweis: Einige Inhalte aus diesem Kapitel sind in Kopp et al. (2022) veröffentlicht.

Betriebliche Einführungsprozesse von Cobots stellen für Unternehmen mitunter eine große Herausforderung dar (vgl. Kapitel 2.3.2). Dies betrifft u. a. die Kalibrierung des Vertrauensniveaus der beteiligten Mitarbeitenden, sodass gleichzeitig deren Nutzungsbereitschaft gewährleistet und eine Überschätzung der Cobot-Fähigkeiten vermieden wird. Der Nachweis des Einflusses von Framing-Prozessen wirft in diesem Zusammenhang ein besonderes Schlaglicht auf interne Kommunikationsstrategien, insbesondere zur Adressierung unerfahrener und unsicherer Nutzer:innen moderner Technologien. Ferner legen die durchgeführten Untersuchungen ein differenziertes Vorgehen nahe, bei dem es einige Fallstricke zu beachten gilt. Tabelle 40 bietet einen Überblick über die wichtigsten praktischen Implikationen für die Unternehmensleitung, die sich auf Basis der herausgearbeiteten wissenschaftlichen Befunde ableiten lassen.

Tabelle 40: Tabellarische Übersicht praktischer Implikationen aus unternehmerischer Perspektive auf Basis verschiedener wissenschaftlicher Befunde.

Wissenschaftlicher Befund	Praktische Implikationen
Relevanz des initialen Vertrauens	<ul style="list-style-type: none"> • Bewusstsein schaffen, dass Vertrauensbildungsprozesse bereits vor der ersten Interaktion beginnen. • Implementierung von Kommunikationsstrategien zur Vertrauenssteigerung bereits vor dem Einführungsprozess.
Einfluss des imaginierten Kontakts	<ul style="list-style-type: none"> • Transparente Informationsweitergabe an Mitarbeitende, damit diese sich bereits frühzeitig mit der Vorstellung der Interaktion vertraut machen können.
Einfluss der Angst vor Arbeitsplatzverlust	<ul style="list-style-type: none"> • Bewusstsein für Auswirkungen und Relevanz dieser Ängste sowie der damit verbundenen relationalen Verortungen schaffen. • Transparente Kommunikation über die Auswirkungen auf die Arbeitsplätze der Mitarbeitenden. • Bewusstsein schaffen, dass Cobots ein Mittel der Kollaboration darstellen, bei dem Mitarbeitende in den Prozess integriert bleiben.
Relevanz der Erwartungskongruenz	<ul style="list-style-type: none"> • Bewusstsein schaffen, dass eine Maximierung des Vertrauens nicht zweckdienlich ist, sondern eine angemessene Kalibrierung anzustreben ist. • Präzise Informationen über Leistungsfähigkeit und Zuverlässigkeit des Cobots einholen und kommunizieren, um Erwartungskongruenz zu gewährleisten.
Einfluss von angemessenem sprachlichen Framing	<ul style="list-style-type: none"> • Vermenschlichende Termini nur verwenden, falls davon auszugehen ist, dass die Mitarbeitenden sich in einer kooperierenden Relation zum Cobot begreifen. Ansonsten stringent maschinelle Umschreibung, allerdings unter Vermeidung des Schemas der perfekten Automation. • Achtsame Kommunikation, die angemessenes Vokabular verwendet und unrealistische Erwartungen an die Leistungsfähigkeit des Cobots vermeidet, um Frustrationen während der Interaktion vorzubeugen. • Kritische Reflexion der Verwendung von anthropomorphen Metaphern.
Freude und Kreativität bei der Benutzung	<ul style="list-style-type: none"> • Positive Effekte der Einführung kommunizieren, wie z. B. ergonomischere Arbeitsweise. • Interesse und Freude an der Benutzung durch die gezielte Schaffung von Experimentierräumen wecken.

Wissenschaftlicher Befund	Praktische Implikationen
Effekt der Darbietungshäufigkeit	<ul style="list-style-type: none"> • Frühzeitig reale Berührungs- und Erfahrungspunkte mit Cobots ermöglichen, um Ängste abzubauen, initiales Vertrauen aufzubauen und gleichzeitig überzogenen Leistungserwartungen entgegenzuwirken.
Individualität der Mitarbeitenden berücksichtigen	<ul style="list-style-type: none"> • Individualität der Mitarbeitenden bei der Planung und Kommunikation hinsichtlich deren Kompetenzüberzeugung, Vorerfahrungen, Grundeinstellung gegenüber Robotern und Ansprüchen an eigene Tätigkeiten anerkennen und einbeziehen. • Kompetenzen der Mitarbeitenden und die Wahrnehmung derselben stärken. • Individuelle Ansprüche an Arbeit, eingebaute Arbeitsabläufe und Veränderungsbereitschaft berücksichtigen. • Mitarbeitende mit Vorerfahrungen und Vorkenntnissen als Pilotnutzende einsetzen.
Dynamische Vertrauensveränderungen	<ul style="list-style-type: none"> • Bewusstsein für die gesteigerte Relevanz der ersten Nutzungserfahrungen in der Phase des vulnerablen Vertrauens schaffen. • Sich verändernde mentale Modelle und Erwartungen an den Cobot über längere Zeit im Blick behalten und insbesondere nach unerwarteten Ereignissen (z. B. Fehler des Cobots) die Interpretation derselben hinterfragen. • Kommunikationsstrategie zur vertrauenssichernden und zutreffend sacherläuternden Erklärung von Fehlern des Cobots (<i>debriefing</i>) implementieren.

Formelle oder informelle Schulungsmaßnahmen im Vorfeld einer Cobot-Einführung zur Kalibrierung des Vertrauens sind ratsam. Aufgrund der unterschiedlichen Voraussetzungen und Einstellungen der Mitarbeitenden erscheint ein *one fit for all*-Ansatz allerdings nicht zweckdienlich. Stattdessen sollten die Schulungsmaßnahmen speziell auf die Schulungsteilnehmenden zugeschnitten werden (Schaefer et al., 2021, S. 285). Das erfordert ein differenziertes Verständnis der individuellen Einstellungen und Befürchtungen der Mitarbeitenden (Meißner et al., 2020). Hierbei sind auch der organisationale Kontext sowie die vorherrschende Unternehmenskultur einzubeziehen (H. Choi & Swanson, 2021, S. 136). Das gilt insbesondere vor dem Hintergrund, dass sich das Mensch-Cobot-Vertrauen bei weitem nicht nur auf den Cobot als technisches Artefakt bezieht, sondern im Arbeitskontext bspw. das Vertrauen in die Unternehmensführung adressiert (vgl. Kapitel 3.4.5). Die Studienergebnisse suggerieren, dass insbesondere die Angst vor einer technologischen Ersetzung ein wesentliches Vertrauenshemmnis darstellt, bei dem der Cobot nur

stellvertretend für menschliche Adressat:innen des Vertrauens auftritt. Das Ziel der Schulungen sollte darin bestehen, zunächst Ängste und Misstrauen abzubauen, ein angepasstes Maß an Vertrauen aufzubauen und den Mitarbeitenden zu ermöglichen, mentale Modelle zu entwickeln, aus denen sich die Fähigkeiten des Cobots möglichst präzise ableiten lassen. Dabei kann es auch nötig sein, durch geeignetes Framing initial überhöhte Erwartungen abzubauen. Dies dient auch der Vorbeugung von starken Vertrauensverletzungen während der Interaktion. Neben der theoretischen Beschäftigung mit der Thematik sind sicherheitsüberwachte praktische Experimentierräume sinnvoll, damit sich die Mitarbeitenden mit dem Cobot vertraut machen und eigene Erfahrungen sammeln können. Dadurch kann sich die Vulnerabilität des Vertrauens zu Beginn der Interaktion im Produktivbetrieb verringern (H. Choi & Swanson, 2021).

In Kapitel 5.1.1 wurde bereits darauf hingewiesen, dass sich Vertrauen während der Interaktion entlang unterschiedlicher Verlaufsmuster dynamisch entwickelt. Während das initiale Vertrauen und frühe Erfahrungen bei einigen Personen die weitere Entwicklung bereits hinreichend stark determinieren können, ist bei anderen die Resilienz des Vertrauens gegenüber (wahrgenommenen) Fehlern des Roboters von herausragender Bedeutung. Entsprechend greift ein Schulungskonzept rein im Vorfeld einer Cobot-Einführung zu kurz, um eine optimale MRI zu gewährleisten. Vielmehr sollten die Erwartungen der Mitarbeitenden in Bezug auf den Cobot kontinuierlich im Blick behalten werden. Besonders nach unerwarteten Fehlern des Cobots ist ein *debriefing* nötig, das transparente Informationen über die Fehlerquelle beinhaltet und die Möglichkeiten des sprachlichen Framings zur Vertrauenskalibrierung nutzt, um damit falschen Schlüssen über die Leistungsfähigkeit des Cobots vorzubeugen (Schaefer et al., 2021, S. 285).

6.1.2 Gesellschaftliche Perspektive

Die neuere wissenschaftliche Literatur zum Vertrauen in MRI akzentuiert die korrekte Kalibrierung des Vertrauens anstatt dessen Maximierung als übergeordnetes Ziel (de Visser et al., 2020; Khavas, Ahmadzadeh & Robinette, 2020; Malle et al., 2020). Die zuvor beschriebenen praktischen Implikationen für Unternehmen formulierten gezielte Maßnahmen zur Vertrauenskalibrierung und verdeutlichten, dass die Anwendung von Methoden zur pauschalen und undifferenzierten Akzeptanzbeschaffung leicht zu einem überhöhten Vertrauen in den Cobot seitens der Mitarbeitenden führen kann, was letztlich den längerfristigen unternehmerischen Interessen zuwiderläuft. Dennoch werden insbesondere soziale Funktionen für Roboter traditionell mit der primären Absicht einer Akzeptanzsteigerung hinzugefügt (Compagna & Muhl, 2012, S. 20; Hindemith, Vollmer, Göpfert, Wiebel-Herboth & Wrede, 2020). Wie die vorhergegangenen Analysen zeigen, stellt das sprachliche Framing ein Mittel dar, um das gesellschaftliche Vertrauen in moderne Technologien zu beeinflussen. Aufgrund der prinzipiellen Interpretationsbedürftigkeit von Sprache und

des starken Gestaltungsspielraums in Bezug auf unbekannte Technologien steht dieses Mittel allen Akteur:innen in der öffentlichen Debatte zur Verfügung. Diese sind imstande, das initiale Vertrauen stark zu prägen, indem sie Technikbeschreibungen sprachlich vermitteln, die letztlich über die Klassifizierung und Bewertung einer Technologie entscheiden. Die Grenzen zwischen einem unvermeidlichen und unbewussten Framing und einer absichtsvollen Manipulation sind dabei fließend. Damit stellt sich die Frage nach einem wünschenswerten und verantwortungsvollen Umgang mit Sprache in Bezug auf Roboter, die auch die Sprachphilosophie und die Technikethik beschäftigt (Coeckelbergh, 2018; Loh, 2019, S. 124f.; Wadepul, 2020).

Die Brisanz dieser Fragestellung ergibt sich aus dem Zusammenspiel zweier parallel wirkender Faktoren. Einerseits ist davon auszugehen, dass in modernen Informationsgesellschaften mehr Informationen vermittelt werden und sich die Anwendungsgebiete für Framing damit ausweiten, z. B. im Rahmen gezielter Überzeugungsversuche im Produkt-Marketing (Stocké, 2002, S. 258). Andererseits kommt gerade in Zeiten tendenziell steigender Ambiguität hybrider Roboter wie Cobots sowie einer zunehmenden Abhängigkeit von Robotern einer realitätsgetreuen Begrenzung des Vertrauens auf ein angemessenes Maß eine hohe Bedeutung zu. Dabei läuft die Entwicklung und Einführung solcher Roboter parallel zur Identifikation eines geeigneten Platzes für diese unbestimmten Entitäten in der Gesellschaft, in der Wirtschaft sowie in der Sphäre des Rechts, der Gefühle und der Moral (Johnson & Verdicchio, 2018). Nur wenn die Fähigkeiten von Robotern realistisch eingeschätzt werden, lässt sich verhindern, dass ein unangemessener Verantwortungsübertrag und eine verminderte Kontrolle dazu führen, dass Fehler von Robotern erhebliche Konsequenzen verursachen (Lyons & Nam, 2021). Massenmediale Darstellungen bspw. der Unterhaltungsindustrie, die die Fähigkeiten von Robotern tendenziell überzeichnen, steigern dabei die Gefahr eines überhöhten Technikvertrauens (Malle et al., 2020, S. 5) sowie eines selbstwertbedrohenden kollektiven Unterlegenheitsgefühls. Letzteres kann besonders dann entstehen, wenn Roboter sprachlich als eine Synthese der jeweils besten Aspekte zweier bis dato kategorial unterschiedlicher Entitäten, nämlich Mensch und Maschine, dargestellt werden. Solche Roboter erscheinen dann gleichzeitig so menschlich, als könnten sie Kolleg:innen als soziale Interaktionspartner:innen ersetzen, und so maschinell, als überträfen sie dank überlegener Fähigkeiten jede menschliche Arbeitsleistung. Wenn Menschen vor dem Hintergrund eines technisierten Menschenbildes nur an den typischen Stärken von Maschinen gemessen werden, erscheinen sie unweigerlich defizitär (Grunwald, 2019a, S. 216ff.). Hinzu kommen damit einhergehende Verwirrungen um den Status von handelndem Subjekt und passivem Objekt, die leicht in unreflektierte Verantwortungszuschreibungen an den Roboter münden können. Diese Verantwortung kann der Roboter allerdings nicht übernehmen. So würde die Übernahme einer Rechenschaftsverantwortung bspw. erfordern, dass der Roboter die Beweggründe seiner Aktionen zu explizieren imstande ist (Höffe, 1997, S. 345). Durch solche Verantwortungszuschreibungen würden die

dahinterstehenden Menschen als Entscheider:innen oder Kontrollinstanz übersehen und exkulpiert, was einer Verantwortungsdiffusion Vorschub leisten würde.¹

Die Gefahr eines unangemessenen Vertrauens besteht insbesondere, da sich der breiten Bevölkerung bis dato nur wenige Möglichkeiten zur realen MRI bieten, die zur Rekalibrierung des Vertrauensniveau auf Basis eigener Erfahrungen führen könnten (Compagna & Muhl, 2012, S. 21). Solange dies nicht der Fall ist, stützen v. a. unerfahrene Personen ihr Vertrauen häufig auf oberflächlich sichtbare Robotereigenschaften und auf durch Dritte vermittelte Informationen, die zu einem überhöhten Vertrauensniveau und damit mittelfristig zu Enttäuschungen und Frustrationen führen können (Malle et al., 2020, S. 3). Zwar stellt die Interaktion mit moderner Technologie für den Großteil der (deutschen) Bevölkerung eine alltägliche Beschäftigung dar, allerdings geschieht dies häufig ohne eine tiefere Einsicht in die innere Architektur und Funktionsweise derselben. Perspektivisch ist von einer vergleichbaren Situation im Umgang mit Robotern auszugehen, was eine unreflektierte Übertragung mentaler Modelle von Menschen auf Roboter begünstigt (Hindemith et al., 2020).

Zur Vorbeugung eines kollektiv überhöhten Robotertrusts werden in der Literatur verschiedene Maßnahmen vorgeschlagen. Einerseits fordern einige Wissenschaftler:innen, Ansätze des ehrlichen Designs (*honest design*) zu implementieren und andererseits die Bildung in Bezug auf moderne Technologien zu stärken (Malle et al., 2020, S. 9). Der Ansatz des ehrlichen Designs fordert eine derartige Robotergestaltung, sodass die Fähigkeiten und v. a. die Unfähigkeiten des Roboters unmittelbar transparent werden, um eine mangelnde Übereinstimmung zwischen mentalen Modellen und realen Fähigkeiten zu vermeiden (Hindemith et al., 2020). Diese Forderung bettet sich in aktuelle Forschungen zu sog. Transparenzinformationen von Robotern und deren Wirkung sowie zu verantwortungsvoller Robotik (*responsible robotics*) ein. Auch in diesem Kontext lautet die zentrale Forderung, dass Roboter so zu gestalten und sprachlich zu framen sind, dass ihre Bedeutung für das menschliche Zusammenleben klar wird und sie in gesellschaftlich wünschenswerter Weise eingesetzt werden können (Johnson & Verdicchio, 2018). Zu diskutieren wäre, inwiefern vermenschlichende Sprache und anthropomorphe Gestaltung dem Gebot des ehrlichen Designs und den Prinzipien der verantwortungsvollen Robotik zuwiderlaufen. Bemerkenswert ist in diesem Zusammenhang, dass Wissenschaftler:innen über ihre technischen Errungenschaften vermenschlichend sprechen (Zijlstra, 2017) und innerhalb der Forschung zur Mensch-Technik-Interaktion häufig auf sog. *wizard of Oz*-Experimente zurückgreifen. Bei diesen Experimenten wird den Versuchspersonen der Eindruck vermittelt, mit einem autonomen System zu interagieren oder zu kommunizieren, das allerdings

¹ Eine Übersicht der populären Positionen der Roboterethik zur zunehmend diskutierten Frage der Verantwortungsübernahme durch Roboter findet sich bspw. in Loh (2019, S. 126ff.) mit einem Fokus auf Verantwortung als relationales Phänomen in der Roboterethik ab S. 192.

in Wirklichkeit unbemerkt von der Versuchsleitung in Echtzeit gesteuert wird (Riek, 2012). Solche Ansätze verletzen streng genommen das Gebot des ehrlichen Designs (Malle et al., 2020, S. 5).

Der Erwerb technischen Wissens – optimalerweise gepaart mit Anreizen zu einer elaborierten, bewussten Verarbeitung neuer Informationen – könnte eine weitere Strategie darstellen, um verzerrten Erwartungen und entsprechenden Vertrauensverletzungen vorzubeugen (Malle et al., 2020). Dies könnte dazu führen, die Empfänglichkeit gegenüber sprachlichem Framing zu verringern und somit eine stärkere Widerstandskraft gegenüber manipulativ eingesetztem Framing aufzubauen (Stocké, 2002, S. 258). Insofern entsteht auch ein bildungspolitischer Auftrag aus der Wirkkraft des Framings, der mit den jüngsten Forderungen der Enquete-Kommission des Bundestages zur Künstlichen Intelligenz in Einklang steht. Diese fordert die verstärkte Integration der Kompetenzvermittlung im Bereich der Informatik in den Schulunterricht und unterstreicht dezidiert die Relevanz verwandter Soft Skills, wie bspw. ein verstärktes Bewusstsein für die gesellschaftlichen Auswirkungen und angrenzende philosophische Fragestellungen (Enquete-Kommission Künstliche Intelligenz, 2020). Gleichwohl sei dabei angemerkt, dass jedes zweckdienliche mentale Modell nicht nur eine abbildende, sondern eine komplexitätsreduzierende Funktion beinhaltet und damit auch bei steigendem Vorwissen mitunter zu durch den Modellcharakter bedingten unzutreffenden Vorhersagen führt (vgl. Kapitel 3.1.2).

Zusammenfassend besteht eine gesellschaftliche Aufgabe in einer reflektierten Sprachverwendung und in der Erschaffung und Etablierung eines angemessenen Vokabulars bezüglich moderner Technologien. Freilich lässt sich Sprachverwendung nicht zentral steuern oder vorschreiben. Umso mehr kommt Forschenden in diesem Themenfeld die Verantwortung zu, eine adäquate Sprachverwendung zu pflegen und ein Bewusstsein für deren Relevanz zu schaffen. Sacherläuternde anthropomorphe Metaphern sollten kritisch in Hinblick auf ungeeignete Vermenschlichungstendenzen reflektiert werden. An dieser Stelle ergeben sich Anschlusspunkte für sprachphilosophische Überlegungen. Hierzu zählt bspw. die Frage, inwiefern Metaphern gar notwendig sind, um die Funktionsweisen und Leistungsmerkmale moderner Technologien zu erklären (Gutmann & Rathgeber, 2010; Nerurkar, Wadephul & Wiegerling, 2019) und inwiefern anthropomorphe Metaphern bei näherer Betrachtung keine (falsch-)deskriptiven, sondern anderweitig nicht vermittelbare pragmatische Aussagen kommunizieren (Damiano & Dumouchel, 2018). Noch grundlegendere Fragestellungen betreffen u. a. hermeneutische Aspekte von techno- und biomorphen Metaphern und deren Bezug zum Wahrheitsbegriff (Gutmann & Knifka, 2015). Welche Wirkkraft sprachliches Framing in dieser gesellschaftlichen Perspektive tatsächlich entfaltet und wie nachhaltig die Effekte sind, lässt sich auf Basis der vorliegenden Dissertation mit dezidiertem Bezug zum Arbeitskontext nur begrenzt beantworten. Die übergreifenden Erkenntnisse legen allerdings die Vermutung nahe, dass die Auswirkungen häufig subtiler und unbewusster Natur sind und sich eher implizit denn explizit niederschlagen. Damit

geht die Gefahr einher, die Konsequenzen unreflektiert verwendeter vermenschlichender Sprache zu unterschätzen. Dies liegt erstens an der Multidimensionalität von Vertrauen, zweitens an dem häufig impliziten Charakter von Anthropomorphisierung, drittens an der hohen Personen- und Kontextabhängigkeit der Framing-Effekte und viertens an dem Facettenreichtum möglicher Mensch-Roboter-Relationen, der in der dualistischen Unterscheidung in Werkzeug oder Teammitglied häufig verkürzt und vernachlässigt wird. Die hier dargestellten Überlegungen verstehen sich als Andeutung von Implikationen, die weiterer Vertiefung und empirischer Fundierung in dem weiter gefassten gesellschaftlichen Kontext bedürfen, der außerhalb des Fokus dieser Dissertation liegt.

6.2 Übertragbarkeit der Ergebnisse

Hinweis: Die Erläuterungen zur Übertragung auf KI-Technologie sind ansatzweise in Kopp (2020) veröffentlicht.

Die vorliegende Dissertation stellt primär Cobots am Arbeitsplatz in den Fokus, erhebt allerdings auch den Anspruch, die Übertragbarkeit der gewonnenen Erkenntnisse auf andere Kontexte darzustellen. Abbildung 6.1 stellt stichpunktartig vier mögliche Zieldomänen einer Übertragung dar. Dazu zählen andere Arten von Robotern, andere moderne Technologien wie KI-Anwendungen, andere Personengruppen und andere Nutzungskontexte außerhalb des industriellen Arbeitsplatzes.

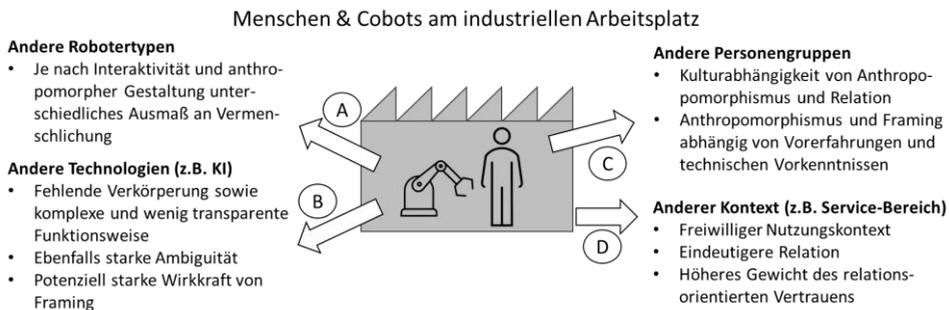


Abbildung 6.1: Übersicht über vier mögliche Zieldomänen (Robotertypen, ähnliche Technologien, Personengruppen, Nutzungskontexte) zur Übertragung der Ergebnisse.

Bereits in der Einleitung dieser Dissertation wurde darauf verwiesen, dass verschiedenartige Typen von (humanoiden und animaloiden) Robotern bereits vereinzelt zu neuen Qualitäten der Mensch-Roboter-Relationen führen und perspektivisch womöglich massenhaft führen werden. Die mangelnde Trennschärfe etablierter Roboter-kategorien und die neuen

Entwicklungen in der Robotik erschweren eindeutige Zuordnungen und damit eine präzise Begrenzung des Anwendungsbereichs der vorliegenden Forschungsergebnisse. Wenn in der vorliegenden Arbeit von Robotern gesprochen wird, sind in aller Regel dezidiert kollaborationsfähige Roboter gemeint. Gleichwohl ist davon auszugehen, dass die Erkenntnisse zum Ablauf von Vertrauensbildungsprozessen in gewissem Umfang auf all diejenigen Roboter übertragbar sind, die eine interaktive Verwendung zu einem gewissen Maße ermöglichen und/oder menschenähnliche Merkmale in der Gestaltung aufweisen (siehe Pfeil A in der Abbildung). Traditionelle Industrieroboter mit physischen Schutzvorrichtungen erfüllen diese Kriterien nicht, da sie insbesondere in einer eindeutigen Werkzeugverwender-Werkzeug-Relation stehen (vgl. Kapitel 2.2.1), wenngleich die Unterschiede und Gemeinsamkeiten in Hinblick auf die Vertrauensbildung bisher kaum dezidiert untersucht wurden (H. Choi & Swanson, 2021, S. 137). Bezugnehmend auf die Abgrenzung in Kapitel 2.1.2 ist zu erwähnen, dass es sich bei der interaktiven Zusammenarbeit nicht notwendigerweise um eine Kollaboration handeln muss. Tendenziell erhöhen eine anthropomorphe Gestaltung und vielfältige Interaktionsmöglichkeiten die Relevanz des Anthropomorphismus, der konstruierten Mensch-Roboter-Relation und des Framings.

Ferner stellt sich die Frage nach der Übertragbarkeit der Ergebnisse auf andere moderne Technologien außerhalb des Feldes der Robotik, die ebenso im Bereich der Arbeitswelt angewendet werden können (Pfeil B in der Abbildung). In der Literatur werden Roboter mitunter mit verwandten modernen Technologien unter weit gefassten Termini wie *intelligente Systeme* (Matthews et al., 2021), *autonome Agenten* (Wynne & Lyons, 2018) oder *intelligente Technologien* im Bereich des Mensch-Maschine-Teamings (Lyons et al., 2019) subsummiert, ohne dass eine Präzisierung vorgenommen wird. Im Rahmen dieses Spektrums an Technologien stellen insbesondere KI-Anwendungen einen interessanten Anwendungsbereich der Erkenntnisse aus dieser Dissertation dar. Diese Anwendungen lassen sich durch ihre fehlende Verkörperung, ihre schwer vermittelbare Funktionsweise und die Intransparenz der Lösungsfindung noch schwieriger mithilfe bekannter mentaler Modelle und mit den Mitteln visueller Darstellungen fassen (Lecheler, 2020). Daraus erwächst mitunter gar der Eindruck, eine Erklärung und Beschreibung solch avancierter Technologien für die breite Masse technischer Laien sei nur noch mithilfe von uneigentlicher Rede wie z. B. anthropomorpher Metaphern möglich (Nerurkar et al., 2019, S. 271). Bereits die Bezeichnung als künstliche Intelligenz stellt einen Rückgriff auf einen Terminus aus der menschlichen Erlebenswelt dar, der wenig zielführend erscheint. Denn für Intelligenz existieren verschiedenste Definitionen (Seng, 2019), sodass sich der KI-Begriff als Projektionsfläche für die unterschiedlichsten Erwartungen darstellt. Das kann leicht zu teils unzutreffenden Assoziationen, zu Kategorienfehlern bei der Zuschreibung von Fähigkeiten (Brooks, 2017) und zu verschwimmenden Grenzen zwischen Subjekt und Objekt führen (Grunwald, 2019b). Öffentliche Debatten thematisieren häufig ferne Zukunftsvisionen in Gestalt spektakulärer und hypothetischer KI-Utopien und -Dystopien (Kehl & Coenen, 2016) oder adressieren hochkomplexe Fragen nach der Intelligenz, Personalität oder

Empfindungsfähigkeit von Robotern (Coeckelbergh, 2009), anstatt aktuelle Leistungsfähigkeiten und -beschränkungen darzustellen. Studien zeigen, dass viele Menschen wenig bewusste Berührungspunkte mit KI und eine sehr unklare Vorstellung von diesen Algorithmen haben, die sie teilweise gar mit Robotik gleichsetzen (Cave, Coughlan & Dihal, 2019; European Commission, 2017; Merenkov, Campa & Dronishinets, 2021). Daher ist von einer hohen Sensitivität für Framing und einem breiten Spektrum möglicher vermenschlicher Zuschreibungen und Relationen auszugehen, sodass die vorliegenden Untersuchungsergebnisse eine hohe Relevanz für diesen Bereich besitzen. Auch die erarbeiteten praktischen Implikationen für innerbetriebliche Technologieadoptionen- und Vertrauensbildungsprozesse scheinen in hohem Maße auf andere Technologieeinführungen übertragbar.

Des Weiteren ist die Anwendbarkeit der Forschungsergebnisse auf verschiedene Personengruppen von Relevanz (Pfeil C in der Abbildung). Da im Rahmen der durchgeführten Experimentalstudien keine für die Gesamtbevölkerung repräsentativen Stichproben untersucht wurden, lässt sich eine Übertragbarkeit auf abweichende Populationen, insbesondere aus anderen Kulturkreisen, nicht belegen (vgl. Kapitel 6.3.2.3 zur eingeschränkten Repräsentativität). Doch selbst innerhalb der deutschen Bevölkerung ist von starken Unterschieden in Hinblick auf relevante Persönlichkeitsmerkmale auszugehen, die sich vermutlich auf die Wirksamkeit von Framing, die Stärke des Anthropomorphismus sowie die Vertrauensbildungsprozesse auswirken. Besonders zu berücksichtigen sind in diesem Zusammenhang primär die subjektive Technikkompetenzüberzeugung, die theoretischen Vorkenntnisse und praktische Vorerfahrungen mit Robotern sowie die individuelle Ambiguitätsaversion und die Disposition zur Vermenschlichung. Unerfahrene Laien sind von technikaffinen und intensiven Techniknutzer:innen zu unterscheiden.

In Hinblick auf den MRI-Kontext ist davon auszugehen, dass die Mechanismen der Vertrauensbildung und der dynamischen Vertrauensentwicklung sowie die Wirkung des Framings und des Anthropomorphismus auch außerhalb des Arbeitsplatzkontextes relevant sind (Pfeil D in der Abbildung). Zu beachten ist allerdings, dass in primär sozialen Interaktionskontexten relationsorientierte Vertrauenskomponenten über eine höhere Relevanz verfügen als leistungsorientierte und damit andere Robotereigenschaften stärker auf die Vertrauensbildung wirken. Ferner entfällt in diesen Situationen die Angst vor der technologischen Ersetzung als charakteristisches Merkmal der Mensch-Roboter-Relation am industriellen Arbeitsplatz. Daher ist anzunehmen, dass die Relationen im Service-Bereich eindeutiger bestimmt werden können und weniger Komplexitätsgrade aufweisen. Ferner handelt es sich um einen freiwilligen Nutzungskontext, in dem Interaktionen bereits ein ausreichend hohes Vertrauensniveau voraussetzen, damit es überhaupt zu einer Nutzungsentscheidung kommt. Insofern ist davon auszugehen, dass das Vertrauensniveau der Techniknutzer:innen in diesem Bereich deutlich höher liegt und sich eher Herausforderungen aufgrund eines überhöhten anstatt eines zu niedrigen Vertrauens ergeben.

6.3 Methodische Reflexion

6.3.1 Eignung des methodischen Ansatzes

Aus methodischer Sicht hat sich der methodenpluralistische und interdisziplinäre Forschungsansatz bewährt. Aufgrund der dargelegten Komplexität und Multidimensionalität des Mensch-Roboter-Vertrauens scheinen multimethodische und interdisziplinäre Betrachtungen notwendig, um ein ausreichend vertieftes Gesamtbild zu erhalten. Der Einsatz quantitativer und qualitativer Forschungsdesigns sowie impliziter und expliziter Erhebungsmethoden führte zu teils konvergierenden und teils komplementären Erkenntnissen. Insbesondere implizite Vermenschlichungstendenzen wären ohne qualitative Ansätze und indirekte Erhebungsmethoden verborgen geblieben. Ferner erwiesen sich die theoretischen Ansätze und Denkfiguren aus der Soziologie, Philosophie und Sprachwissenschaft insbesondere für die Interpretation und Kontextualisierung der empirischen Ergebnisse als gewinnbringend. Die deutlichen Unterschiede zwischen den beiden Populationen und den jeweiligen Anwendungssettings unterstreichen außerdem die Relevanz einer anwendungsorientierten Forschung, die die externe Validität in den Vordergrund rückt. Tabelle 41 liefert einen Überblick über die Qualitätsmerkmale des Untersuchungsdesigns.

Tabelle 41: Übersicht über Qualitätsmerkmale des Untersuchungsdesigns.

Qualitätsmerkmale des Untersuchungsdesigns	Erläuterung
Methodenpluralismus	Verwendung qualitativer und quantitativer Forschungsansätze, um ein ganzheitliches Bild zu erhalten.
Interdisziplinarität	Kombination von Theorieansätzen, in erster Linie aus Psychologie und Kognitionswissenschaft (z. B. Drei-Faktoren-Theorie des Anthropomorphismus), Soziologie (z. B. Vertrauenskonzeption nach Luhmann, Modell der Frame-Selektion) und Philosophie (z. B. Roboter als Quasi-Gegenüber, relationale Wende der Ethik, intentionale Grundhaltung).
Hohe interne Validität (Experimentalstudien)	Aktive experimentelle Manipulation, randomisierte Gruppenzuteilung in beiden Experimentalstudien, Kontrollgruppe in Experimentalstudie 1, doppelt-blindes Design in Experimentalstudie 2.
Hohe externe Validität (Vorstudien und Experimentalstudie 1)	Produktionsmitarbeitende zur Gewährleistung einer hohen ökologischen Validität; mind. 25 Versuchsteilnehmende pro Versuchsgruppe.

Qualitätsmerkmale des Untersuchungsdesigns	Erläuterung
Hohe externe Validität durch realitätsnahes Setting (Vorstudie 2 und Experimentalstudie 2)	Interviews und Workshops in Vorstudie 2 in Unternehmensumgebung; Experimentalstudie 2 in Präsenz-Setting (in Laborumgebung) mit realistischer MRI mit kollaborierenden Anteilen und realem Cobot.
Verschiedene Erhebungsmethoden (Experimentalstudien)	Selbstauskunftsskalen und indirekte Messmethoden (quantitative Durchdringung von Textmaterial) zur Erhebung von implizitem Anthropomorphismus.
Validierte Multi-Item-Skalen (Experimentalstudien)	Vorwiegend validierte und vielfach verwendete Multi-Item-Skalen mit hoher interner Konsistenz.

Die gleichwohl existierenden methodischen Einschränkungen werden in Kapitel 6.3.2 erläutert. Auf Basis dieser Limitationen sowie der ausführlichen Einbettung der empirischen Ergebnisse dieser Dissertation in die aktuelle Forschungslandschaft in Kapitel 5 ergeben sich sinnvolle inhaltliche Stoßrichtungen sowie methodische Anforderungen für anknüpfende Forschung, die in Kapitel 6.4 skizziert werden.

6.3.2 Methodische Einschränkungen

6.3.2.1 Begrenzte externe Validität durch experimentelles Setting

Trotz des Bemühens um ein möglichst anwendungsnahe Forschungsdesign konnte im Rahmen der vorliegenden Dissertation aufgrund forschungspragmatischer Einschränkungen und der zusätzlich erschwerten Rahmenbedingungen im Zuge der Covid-19-Pandemie kein Experiment durchgeführt werden, bei dem Produktionsmitarbeitende mit einem realen Cobot interagierten. Insofern decken die beiden Experimentalstudien jeweils einen von zwei entscheidenden Faktoren für die Übertragbarkeit in die betriebliche Praxis ab, nämlich den Rückgriff auf die entsprechende Stichprobe bestehend aus Produktionsmitarbeitenden (Experimentalstudie 1) und die Durchführung einer realitätsnahen MRI (Experimentalstudie 2) – wenngleich in einer Laborumgebung. Die höchste externe Validität ergäbe sich gleichwohl bei der Kombination beider Faktoren in einem Experiment innerhalb der realen Arbeitsumgebung. Nichtsdestotrotz konnte durch das gewählte Design eine externe Validität erzielt werden, die das Gros der bisher durchgeführten Studien übertrifft, da diese zumeist Studierende als Stichprobe heranziehen und keine reale Interaktion beinhalten (vgl. Kapitel 4.1).

6.3.2.2 Längerfristige Vertrauensentwicklungen und Framing-Effekte

Bei einer realen betrieblichen Cobot-Einführung ist die Vertrauensentwicklung über einen deutlich größeren Zeitraum als in den Experimentalstudien relevant. Die empirischen, fallstudienbezogenen Erkenntnisse von Wurhofer et al. (2015) verdeutlichen, dass dieser Prozess in verschiedenen Phasen verläuft, die durch unterschiedliche Facetten gekennzeichnet sind (vgl. Kapitel 2.3.2.1). Auch das Drei-Phasen-Modell des Anthropomorphismus verweist auf zeitlich unterschiedliche Muster (vgl. Kapitel 3.2.4.2). Zwar wurde Vertrauen zu verschiedenen Zeitpunkten gemessen, allerdings lässt sich damit die Dynamik der Vertrauensentwicklung nur verkürzt und diskretisiert abbilden. Die einzelnen Messzeitpunkte stellen insofern zeitpunktbezogene Einblicke in einen zeitkontinuierlich verlaufenden Prozess dar. Entsprechend lässt sich auch die Effektivität von Framing-Interventionen nicht in Bezug auf einen länger währenden Nutzungszeitraum, sondern gemäß dem *ex ante* definierten Fokus der vorliegenden Dissertation nur in Bezug auf das initiale vermittelte Vertrauen festlegen. In der Praxis mit typischerweise länger andauernden Interaktionen sind gleichwohl die längerfristigen Auswirkungen im Zusammenspiel mit zusätzlichen Einflussfaktoren durch die Erfahrungen und Beobachtungen der stattfindenden MRI relevant.

6.3.2.3 Eingeschränkte Repräsentativität

Die Versuchsteilnehmenden in den durchgeführten Studien bildeten jeweils keine repräsentative Stichprobe ab. Da sich die Stichproben ferner aus in Deutschland lebenden und der deutschen Sprache mächtigen Personen zusammensetzten, besteht in der Verallgemeinerbarkeit auf andere Staatsangehörige oder Angehörige anderer Kulturkreise eine weitere Limitation. Einige vorige Studien haben bereits darauf hingewiesen, dass kulturelle Modelle (*cultural models*) die Wahrnehmung, die Grundeinstellung und die Art und Weise der Interaktion mit Robotern beeinflussen (Bartneck, Kanda, Mubin & Al Mahmud, 2009; Mutlu et al., 2016, S. 1910). So betrachteten koreanische und japanische im Vergleich zu deutschen Versuchsteilnehmenden soziale Roboter als vertrauenswürdiger (Li, Rau & Li, 2010). Länderspezifische Auswirkungen betreffen auch die Stärke des Anthropomorphismus, der laut einer nicht-repräsentativen Studie z. B. bei Niederländern signifikant stärker vertreten zu sein scheint als bei Deutschen (de Graaf & Ben Allouch, 2013). Insofern können die empirischen Studienergebnisse nicht ohne Weiteres auf andere Nationalitäten und Kulturkreise übertragen werden. Dennoch ist davon auszugehen, dass sich die grundlegenden kognitiven Mechanismen ähnlich darstellen, zumal personenindividuelle Unterschiede tendenziell stärker ins Gewicht fallen als nationale (Turja & Oksanen, 2019).

6.3.2.4 Messung von Vertrauen und Misstrauen als singuläre Konstrukte

In der vorliegenden Dissertation wurden Vertrauen und Misstrauen jeweils als singuläre Konstrukte gemessen. Die einzelnen Items der TAS wurden dabei zu einem aggregierten Vertrauens- sowie Misstrauenswert zusammengefasst, der jeweils alle Dimensionen des

Vertrauens und des Misstrauens beinhaltet. Für eine weitere Ausdifferenzierung ist die TAS nicht vorgesehen. Ein Ergebnis der Dissertation stellt allerdings die Erkenntnis dar, dass Vertrauen und Misstrauen auch oder insbesondere am Arbeitsplatz multidimensionale Konstrukte darstellen. Die einzelnen Dimensionen lassen sich allerdings messtheoretisch mit der TAS nur unzureichend erfassen. Eine genauere Unterscheidung der Bezugsobjekte des Vertrauens hätte potenziell in höherem Maße ausdifferenzierte und tiefere Einblicke ermöglicht. Hierzu wären allerdings multidimensionale Messinstrumente notwendig gewesen, die allerdings erst nach der Durchführung der empirischen Studien entwickelt und etabliert wurden (vgl. Kapitel 5.1.3 zu multidimensionalen Vertrauenskonzepten).

6.3.2.5 Bewertung der Interaktion

Analog zur Manipulationskontrolle sollte bei Experimentalstudie 2 die Güte der Interaktion – abgesehen von der quantitativen Selbsteinschätzung – ergänzend durch die qualitative Auswertung des Verhaltens der Versuchsteilnehmenden während der Interaktion mittels der Analyse entsprechender Videoaufnahmen erhoben werden. Damit hätte neben den quantitativen Selbsteinschätzungen zur Interaktion ein zweites verhaltensorientiertes Maß zur Verfügung gestanden, was die Reliabilität der Auswertungen erhöht hätte. Allerdings erwies sich das Videomaterial als nicht reichhaltig und eindeutig genug, um zuverlässig Rückschlüsse abzuleiten, wie sich die Studierenden während der Interaktion fühlten und wie sie den Cobot wahrnahmen. So konnten zwei unabhängige Gutachter bei der Kodierung des Videomaterials anhand eines eigens entwickelten Kodierleitfadens keine ausreichend hohe Übereinstimmung erzielen. Daher wurde dieser Ansatz verworfen. Elaboriertere objektive Messinstrumentarien wie *eye tracking*, *motion tracking* oder Stimmanalysen hätten möglicherweise zuverlässige Rückschlüsse erlaubt. Der Einsatz solcher aussichtsreicher verhaltensorientierter Messmethoden wird derzeit in der Forschungsgemeinschaft diskutiert (Schaefer et al., 2021, S. 278).

6.3.2.6 Operationalisierung der Art der Mensch-Cobot-Relation

Ein erklärtes Ziel dieser Dissertation bestand darin, die Mensch-Cobot-Relation als möglichen Einflussfaktor auf das Mensch-Cobot-Vertrauen einzubeziehen. Im Gegensatz zur wahrgenommenen Menschenähnlichkeit, die bereits in vielen bestehenden Studien als UV verwendet wurde, standen für die Operationalisierung der Mensch-Cobot-Relation keine erprobten und validierten Instrumente zur Verfügung. Am nächsten verwandt erscheint der Ansatz von Hinds et al. (2004), die den „(relativen) Status“ (*relative status*) des Roboters manipulierten, dabei allerdings nicht auf striktes Framing, sondern auf gezielte Informationsvariation zurückgriffen. Der Roboter wurde entweder als Vorgesetzte:r (*supervisor*) oder als Untergebene:r (*subordinate*) bzw. als gleichrangig (*peer*) vorgestellt. Die Manipulation bezog sich demnach auf formalisierte Hierarchiegefüge und nicht auf die Qualitäten einer Mensch-Cobot-Relation. Insofern kam in Experimentalstudie 1 eine eigens entwickelte Ad-hoc-Skala zum Einsatz, die zwar eine gute interne Konsistenz aufwies

(Cronbachs $\alpha = .854$), aber nichtsdestotrotz einer Validierung mittels methodologischer Untersuchungen bedarf (Döring & Bortz, 2016, S. 440ff.).

Eine weitere Herausforderung bei der Entwicklung einer Operationalisierung der Mensch-Cobot-Relation stellte die Gewährleistung einer ausreichenden Diskriminanzvalidität dar. So zeigte sich in Experimentalstudie 1 eine signifikant schwach positive Korrelation mit der wahrgenommenen Menschenähnlichkeit, $r = .207$, $p = .011$. Da die beiden Texte zur experimentellen Manipulation der UVs nacheinander dargeboten wurden, bevor die Manipulationskontrolle für beide UVs folgte, könnte ein durch den Versuchsablauf bedingter *carry-over*-Effekt die statistisch nachweisbare Korrelation verursacht haben. Aufgrund des niedrigen Korrelationskoeffizienten wird allerdings keine Kollinearität angenommen. In Folgeexperimenten könnten Distraktoraufgaben eingeführt und die jeweiligen Manipulationskontrollen direkt nach den entsprechenden Texten durchgeführt werden, um solche Effekte zu minimieren.

6.3.2.7 Identische Informationen beim sprachlichen Framing

In Kapitel 5.4.1 wurde darauf hingewiesen, dass im Rahmen dieser Dissertation Framing im strikten Sinne verstanden wird. Das bedeutet, dass Unterschiede in der Art der Informationsvermittlung, nicht aber in den Informationen selbst als ursächlich für die Aktivierung unterschiedlicher Schemata angesehen werden. Während zur Variation der Texte zur Menschenähnlichkeit einige Beispiele aus der Literatur herangezogen werden konnten, war dies in Bezug auf die Mensch-Cobot-Relation nicht der Fall. Die Eignung der gewählten Variationen lässt sich daher in Frage stellen. Ferner ist fraglich, inwiefern sich die Identität der Informationen bei einem solch anwendungsnahen Forschungsgegenstand überhaupt gewährleisten lässt. Aufgrund dessen werden Studien mit Hervorhebungsframes durchaus methodisch kontrovers diskutiert (Cacciatore, Scheufele & Iyengar, 2016). In diesem Ansatz bestand ein Unterschied zu bekannten Studien aus der Entscheidungstheorie. In Letzteren werden im Rahmen der experimentellen Manipulation nur gezielt einzelne Begriffe oder komplementäre Wahrscheinlichkeitswerte ausgetauscht, wie bspw. beim bekannten *Asian disease problem* (Stocké, 2002; Tversky & Kahneman, 1981). Das Vorgehen in der vorliegenden Dissertation fokussierte die grundsätzliche Wirksamkeit von Framing im MRI-Anwendungskontext und griff damit auf Formulierungsänderungen zurück, die denselben Informationsgehalt aufweisen sollten. Der Anteil von einzelnen Variationen am Gesamteffekt des Framings lässt sich mit diesem Ansatz im Nachhinein nicht isoliert quantifizieren, sondern wäre ein geeigneter Gegenstand anknüpfender und stärker theoriebildender Forschungsarbeiten.

6.4 Ausblick auf anknüpfende Forschung

6.4.1 Relationale Perspektive und Anthropomorphismus

Neben der intensiven Erforschung der ontologisch orientierten Frage nach dem Wesen und den mentalen Zuständen von Robotern besteht weiterer Forschungsbedarf aus einer phänomenologisch-relationalen Perspektive auf Cobots. Das betrifft konkret die Frage, wie sich Menschen relational zu Cobots verorten und welche Konsequenzen das hat. Die Angst vor Arbeitsplatzverlust betrifft neben der Wahrnehmung des Cobots auch die mentale Konzeption der Beziehung zu demselben, die sich wiederum im Zusammenspiel mit der wahrgenommenen Menschenähnlichkeit auf das Vertrauen auswirkt. Dennoch bleibt dieser Faktor in den allermeisten Studien unberücksichtigt und mindert die Anwendbarkeit der Studienergebnisse im Arbeitskontext. Insbesondere das komplexe und kontextspezifische Wechselspiel zwischen der Mensch-Cobot-Relation und anderen Faktoren wie dem Anthropomorphismus bedarf einer ausgiebigen Untersuchung (vgl. Kapitel 5.3.2). Ferner sollte untersucht werden, inwiefern aktuelle philosophische Strömungen das Verständnis der Mensch-Cobot-Relation vertiefen und weiterentwickeln können (vgl. Kapitel 5.2.2)

Die Operationalisierung der Mensch-Cobot-Relation sollte über tradierte Dichotomien aus der Debatte um Teammitglied oder Werkzeug hinausgehen und die vielschichtigen Ebenen möglicher Beziehungsformen sowie damit verbundene Erwartungshaltungen und Wechselwirkungen explorieren. Neben der im Rahmen der vorliegenden Dissertation verwendeten Dimension einer kooperierenden im Gegensatz zu einer konkurrierenden Relation erscheint der Einbezug weiterer qualitativer Dimensionen ratsam (vgl. Kapitel 5.2.3). In diesem Zusammenhang könnte die Integration von Persönlichkeitsmerkmalen als mögliche Kovariaten die Aussagekraft der Ergebnisse erhöhen, zumal frühere Studien Zusammenhänge zwischen *big five*-Persönlichkeitsmerkmalen und der Einstellung gegenüber Robotern vermuten lassen (Müller & Richert, 2018).

6.4.2 Kontextspezifische und zeitdynamische Wirkbedingungen von Framing

Weitere Anknüpfungspunkte für Forschungsvorhaben bestehen in Hinblick auf die kontextspezifischen Wirkbedingungen von Framing. Im Rahmen der vorliegenden Dissertation wurden theoretische Ansätze zur Erklärung der gefundenen Framing-Effekte vorgeschlagen (vgl. Kapitel 5.4.2). Weitere methodisch stringente empirische Studien sind notwendig, um zu ergründen, welche Bedingungen die Wirkkraft von striktem Framing beeinflussen. Das betrifft sowohl personenspezifische als auch situative Faktoren sowie Merkmale der Framing-Texte. In diesem Zusammenhang sollte bspw. geklärt werden, wie

sich konkrete Variationen in der Informationsvermittlung auf verschiedene Facetten des Vertrauens auswirken und welche Erwartungen diese hervorrufen. Ferner sollte die insbesondere aus praktischer Sicht relevante Frage nach der Persistenz von Framing-Effekten im Verlauf länger andauernder Interaktionen sowie über aufeinanderfolgende Interaktionen hinweg untersucht werden (vgl. Kapitel 5.4.4). Die im Rahmen dieser Dissertation entwickelten Ansätze können als Startpunkte für eine empirisch getriebene Theoriebildung dienen, die sich der bestehenden Evidenz aus klassischen Anwendungsgebieten des Framings bedient.

6.4.3 Replikation der Experimentalstudien

Eine naheliegende und konkrete Anknüpfungsmöglichkeit stellt die Replikation der Experimentalstudien unter Einbezug einer höheren Stichprobenanzahl dar. Auf Basis der vorliegenden Daten lässt sich nicht klären, ob die teils unterschiedlichen Befunde aus beiden Experimentalstudien in Hinblick auf die Wirkkraft des Framings entweder messtheoretisch durch die Stichprobenanzahl, durch die unterschiedlichen Populationen oder durch die verschiedenen Settings (*online* vs. *offline*) bedingt waren. Zur weiteren Erhöhung der externen Validität böte es sich daher an, Experimentalstudie 2 mit Produktionsmitarbeitenden durchzuführen. Hierdurch ließe sich auch empirisch abklären, ob die Merkmale der Stichprobe oder das experimentelle Setting die Wirkkraft des Framings beeinflussten. Alternativ wäre zu diesem Zweck auch eine weniger aufwändige Replikation der Experimentalstudie 1 mit Studierenden möglich. Im Rahmen dieser Replikationsstudien könnte bspw. der implizite Assoziationstest (*implicit association test*, IAT) als verhaltensbezogenes Maß ein weiteres attraktives Messinstrument darstellen (vgl. Wullenkord & Eyssel, 2019), um bspw. zu bestimmen, welche Assoziationen und Erwartungen die Versuchsteilnehmenden mit unterschiedlich beschriebenen Robotern verbinden und inwiefern sie ihre Relation zu diesen als kooperierend oder konkurrierend repräsentieren. Da sich die Rede über Roboter als aussagekräftige Quelle für Rückschlüsse auf die zugrunde liegenden mentalen Modelle erwiesen hat, ließe sich ferner bspw. die *Methode des (periaktionalen) lauten Denkens* (Döring & Bortz, 2016, S. 371) einsetzen, um zusätzliches sprachliches Material zu generieren, dieses einer strukturierten Textanalyse zu unterziehen und damit vertiefte und qualitative Einblicke in die aktivierten mentalen Modelle zu erhalten (Preim & Dachselt, 2010, S. 96).

6.4.4 Differenzierte multidimensionale Vertrauensbetrachtung

Weiterer Forschungsbedarf besteht in einer vertieften Betrachtung und stärkeren Ausdifferenzierung von Vertrauen am Arbeitsplatz unter Berücksichtigung des jeweiligen

Kontexts. Die meisten Studien messen Vertrauen als singuläres Konstrukt mittels einer aggregierten Skala. So bezieht sich die viel zitierte Studie von Waytz et al. (2014) zum Zusammenhang von Anthropomorphismus und Vertrauen explizit nur auf das Vertrauen in die Leistungsfähigkeit der Maschine. Wie die Erfahrungen und Erkenntnisse aus der vorliegenden Dissertation nahelegen, stellt dies eine simplifizierte Sicht auf Vertrauen – gerade im Kontext der Roboternutzung am Arbeitsplatz – dar. Besonders in diesem Kontext sollte stets die Frage gestellt werden, *worauf* genau vertraut wird und welche Dimensionen des Vertrauens dabei relevant sind (vgl. Kapitel 5.1.3). Ferner sollten gängige Modelle Vertrauen nicht als Merkmal einer dyadischen Relation zwischen einer Person und einem Cobot betrachten, sondern den Kontext einbeziehen, zu dem weitere Adressat:innen des Vertrauens sowie die spezifische Mensch-Cobot-Relation gehören. Im Zusammenhang mit den relevanten Vertrauensdimensionen sind überdies weitere experimentelle Studien notwendig, die die in den Experimentalstudien andeutungsweise vorhandenen Unterschiede und Wechselwirkungen zwischen Vertrauen und Misstrauen gezielt untersuchen und weiter ergründen (vgl. Kapitel 5.1.4)

6.4.5 Ganzheitliche und zeitdynamische Vertrauensmodelle

Weiterer Forschungsbedarf besteht in Hinblick auf zeitdynamische Veränderungsprozesse des Mensch-Roboter-Vertrauens. Die Vielfalt der Effekte auf die zeitlich dynamische Vertrauensentwicklung und deren Zusammenspiel ist bisher weder umfassend verstanden noch empirisch untersucht. Zur näheren Bestimmung von Vertrauensverläufen sind Studiendesigns mit mehreren Messzeitpunkten oder mit kontinuierlichen Messmethoden wie bspw. strukturierten Verhaltensbeobachtungen oder physiologischen Messmethoden wie *eye tracking* ratsam. Gleichwohl erlaubt ein Studiendesign mit mehreren Messzeitpunkten nur ein künstlich diskretisiertes Abbild des Vertrauensverlaufs während einer MRI. Quantitative und qualitative Erhebungsmethoden, die auf kontinuierlicher Verhaltensbeobachtung basieren, bieten zwar eine bessere zeitliche Auflösung, ermöglichen allerdings zunächst nur Rückschlüsse auf verhaltensbezogene Vertrauenskomponenten. Insbesondere in einem verpflichtenden Nutzungskontext, in dem externe Faktoren wie die Anweisungen einer Führungskraft die Handlungen maßgeblich beeinflussen, ist allerdings zu bezweifeln, dass sich das empfundene Vertrauen im Sinne einer Einstellung aus den Handlungen ableiten lässt (Meißner et al., 2020).

Wenn eine längerfristige Beobachtung einer MRI aus forschungspragmatischen Gründen unmöglich ist, kann alternativ auf longitudinale Studien zurückgegriffen werden, die auf wiederholten, jeweils kurzzeitigen Interaktionen zwischen Mensch und Roboter im Labor basieren (Malle et al., 2020, S. 15). In diesem Zusammenhang stellt das zeitpunktabhängige Zusammenspiel zwischen Bottom-Up- und Top-Down-Faktoren ein aussichtsreiches

Forschungsgebiet dar. Ferner bietet sich auch die Untersuchung der Wirkmechanismen des imaginierten und realen Roboterkontakts an. Bisher werden diese Effekte zwar im MRI-Kontext als Erklärungsmuster ins Spiel gebracht, es bedarf allerdings einer besseren theoretischen Absicherung und einer experimentellen Untersuchung in verschiedenen Kontexten und unter Einbezug der subjektiven Wahrnehmung der MRI. Das Zielbild bestünde in einem ganzheitlichen und praxisnah einsetzbaren Modell der dynamischen Vertrauensentwicklung über längere Zeitabschnitte hinweg und unter Explikation der entsprechenden Einflussfaktoren, wie in Kapitel 5.1.1 schematisch skizziert.

Daran schließt sich auch die grundsätzliche Empfehlung an, insbesondere die Integration bestehender empirischer und teils konfligierender Evidenzen zu Einzelphänomenen in ganzheitliche und kontextsensitive Theoriegebilde voranzutreiben. Auch umfassende Taxonomien, die sich in realitätsnahen Anwendungsszenarien außerhalb der sozialen Robotik anwenden lassen (siehe hierzu Onnasch & Roesler, 2021), können einen wichtigen Beitrag zur Konsolidierung einzelner Studienergebnisse leisten. Letztere sind bisher aufgrund uneinheitlicher oder unklarer Definitionen der zentralen Konstrukte wie Vertrauen, Akzeptanz oder Framing sowie der Verwendung unterschiedlicher Roboter und Interaktionskontexte nur bedingt vergleichbar.

7 Zusammenfassung und Fazit

Die vorliegende Dissertation verfolgte primär das Ziel, das Mensch-Cobot-Vertrauen am Arbeitsplatz und dessen Beeinflussbarkeit durch sprachliches Framing zu untersuchen. Die Motivation für diesen Anwendungskontext ergab sich aus der Tatsache, dass Cobots durch die vielfältigen Interaktionsmöglichkeiten und ihre mitunter anthropomorphe Gestaltung Fragen des Mensch-Technik-Vertrauens besonders pointiert stellen und bereits heutzutage Berührungspunkte zwischen Menschen und Robotern am Arbeitsplatz schaffen. Für diesen Untersuchungsgegenstand eignete sich eine anwendungsnahe Forschung, aus der sich sowohl unternehmerische als auch gesellschaftliche Implikationen und mögliche Übertragungen auf Vertrauensbildungsprozesse in Bezug auf ähnliche Technologien ergeben. Die Dissertation verfolgte dabei einen methodenpluralistischen und interdisziplinären empirischen Ansatz, der zwei Vorstudien und zwei Experimentalstudien umfasste.

Zunächst legte das theoretische Kapitel zur industriellen MRI durch die Herausarbeitung der verschiedenen Interaktionsformen und verfügbaren Robotertypen sowie der Spezifika des industriellen Nutzungskontexts den Grundstein für die weiterführende Untersuchung. Ein Schwerpunkt lag hierbei auf der Beschreibung von Cobots als hybride Roboter, die verschiedenartig mental repräsentiert werden. Das anschließende Kapitel zur kognitiven MRI fokussierte die relevanten kognitiven Konstrukte und deren Verhältnis zueinander. Mentale Modelle dienen als kognitive Repräsentationsform von Cobots und von entsprechenden Interaktionsszenarien. Der Begriff des Anthropomorphismus umschreibt die menschliche Tendenz, technologischen Artefakten wie Robotern menschliche Eigenschaften zuzuschreiben, d. h. situativ in Bezug auf Roboter mentale Modelle zu aktivieren, die menschliche Eigenschaften beinhalten. Sprachliches Framing als ein Prozess gezielter Informationsvermittlung kann solche vermenschlichenden Zuschreibungen provozieren und wirft damit ein Schlaglicht auf die vielgestaltige Rolle von Sprache im Prozess der Deutung moderner Technologien. All diese wechselwirkenden kognitiven Prozesse beeinflussen und prägen letztlich das Mensch-Roboter-Vertrauen.

Viele Forschende aus unterschiedlichen Disziplinen haben sich bereits mit dem Technikvertrauen beschäftigt. Dennoch existiert bis dato keine einheitliche Definition für Vertrauen in diesem Kontext. In Anlehnung an das soziologische Vertrauenskonzept nach Luhmann (1968/2014) und die stark rezipierte Vertrauensdefinition für den Automationsbereich nach J. D. Lee und See (2004) wurde Vertrauen einstellungsorientiert definiert. Demnach lässt sich Vertrauen als eine mentale komplexitätsreduzierende Festlegung auf die Zukunftserwartung verstehen, dass ein betreffender Roboter in einer Situation von Vulnerabilität und Unsicherheit bei der Erreichung eigener Ziele behilflich sein wird.

Misstrauen ist hingegen als alternativer Mechanismus zur Komplexitätsreduktion durch die gegenteilige Erwartung einer negativen Auswirkung auf die Erreichung eigener Ziele gekennzeichnet. Ein eigens entwickeltes integriertes Vertrauensmodell führte bestehende Ansätze wie das Drei-Schichten-Modell nach Hoff und Bashir (2015) und das Drei-Faktoren-Modell nach Hancock, Billings und Schaefer, et al. (2011) zusammen und verdeutlichte, dass eine angemessene Vertrauenskalibrierung anstatt einer -maximierung das Gelingen einer MRI begünstigt.

Die durchgeführten Vorstudien dienten in erster Linie der Klärung der praktischen Relevanz des Mensch-Roboter-Vertrauens für den Erfolg einer betrieblichen Cobot-Einführung und -Nutzung (*FFI*). Mittels sowohl quantitativer als auch qualitativer Erhebungsinstrumente wurden die jeweiligen Einschätzungen von Produktionsmitarbeitenden in leitender und in operativer Funktion erhoben. Dabei maßen die Untersuchungsteilnehmenden dem Vorhandensein von Vertrauen sowohl im Vorfeld einer Cobot-Einführung (initiales Vertrauen) als auch während einer Interaktion (dynamisches Vertrauen) eine hohe Bedeutung bei. In einer frühen Phase der Cobot-Einführung erachteten sie dieses als besonders vulnerabel, sodass erwartungsinkongruente Beobachtungen, wie bspw. einer mangelnden Zuverlässigkeit des Cobots, zu starken Vertrauensrückgängen führen können. Der viel beforschte Einfluss von Vermenschlichung wurde hingegen in expliziten Abfragen als weniger wichtig erachtet. Gleichwohl zeichneten sich in qualitativen Aussagen mitunter implizite Formen schwacher Vermenschlichung ab, die sich vor dem Hintergrund der Theorie des dualen Anthropomorphismus verstehen lassen. Besonders stark zum Vorschein traten ferner existenzielle Ängste in Hinblick auf eine technologische Ersetzung und den Verlust des eigenen Arbeitsplatzes. Diese Bedenken stellen eine Facette des Technikvertrauens dar, die am industriellen Arbeitsplatz und explizit in Bezug auf Roboter jeder Art traditionell eine dominante Rolle spielt und auch die weiterführende Experimentalforschung prägte. Die mit dieser empfundenen Bedrohung einhergehende Verunsicherung der Mitarbeitenden offenbarte sich in einer hohen Aufmerksamkeit und Sensibilität gegenüber (vermeintlichen) Hinweisen auf den geplanten Einsatzzweck und relativen Status des Cobots. Neben der zeitdynamischen Entwicklung entlang verschiedener Phasen ist damit bereits ein zweites Charakteristikum des Vertrauens am Arbeitsplatz benannt, nämlich die Existenz multidimensionaler Bezugsobjekte, die u. a. die Arbeitssicherheit, die Angst vor Arbeitsplatzverlust sowie vor sozialer Isolation oder die Wertschätzung als Mitarbeitende betreffen.

Die folgenden Experimentalstudien griffen diese Erkenntnisse auf, vertieften das Verständnis des Mensch-Cobot-Vertrauens und legten einen verstärkten Fokus auf dessen Einflussfaktoren. Die erste Experimentalstudie untersuchte in einem Online-Setting gezielt das initiale Vertrauen von Produktionsmitarbeitenden, um eine hohe externe Validität zu gewährleisten. Eine randomisierte Gruppenzuweisung, eine aktive experimentelle Manipulation, eine Prä-Post-Messung sowie die Hinzunahme einer Kontrollgruppe entsprachen

dem Goldstandard für Versuchsdesigns zur Erzielung einer hohen internen Validität. In einem ersten Schritt ließ sich experimentell nachweisen, dass sich die wahrgenommene Menschenähnlichkeit eines Cobots und die wahrgenommene Kooperationsorientierung einer Mensch-Cobot-Relation durch die Lektüre eines sprachlich geframten einführenden Textes über einen Cobot verändern ließ (*FF2*). Erwartungskonform führte eine Beschreibung mit Termini aus dem menschlichen Begriffsraum zu einer stärker menschenähnlichen Wahrnehmung. Ebenso hatte eine Beschreibung mit einem starken Fokus auf der unterstützenden Wirkung des Cobots in der Zusammenarbeit mit dem Menschen eine kooperierend konstruierte Mensch-Cobot-Relation zur Folge. Der explizite Einbezug der Mensch-Cobot-Relation als experimentelle Variable stellte dabei ein Novum in der MRI-Forschung dar. Dieser Einbezug war durch die in den Vorstudien sichtbar gewordene Relevanz der Angst vor Arbeitsplatzverlust und der damit verbundenen ambivalenten Haltung gegenüber Cobots motiviert, die sich aus dem Spannungsfeld zwischen Unterstützung und Ergänzung auf der einen sowie Unterlegenheit und Ersetzung auf der anderen Seite speist.

In einem zweiten Auswertungsschritt ließ sich zeigen, dass die beiden Faktoren der wahrgenommenen Menschenähnlichkeit und der wahrgenommenen Kooperationsorientierung im Zusammenspiel das initiale Vertrauen von Produktionsmitarbeitenden in einen Cobot beeinflussten (*FF3*). Die Mensch-Cobot-Relation moderierte die Effekte der Vermenschlichung auf Vertrauen. Dieser Interaktionseffekt lässt sich anhand der verschiedenen relevanten Vertrauensdimensionen, der Drei-Faktoren-Theorie des Anthropomorphismus und der differenzierten psychischen Reaktionen auf antizipierte oder reale technologische Ersetzung erklären. Nur wenn sich Mitarbeitende in einer kooperierenden Relation zu einem Cobot verorten, wirkt eine starke Vermenschlichung positiv auf das empfundene Vertrauen. In einer konkurrierenden Relation unterminiert möglicherweise die Angst vor der Ersetzung das Vertrauen. Gemäß einer aktuellen Studie von Granulo et al. (2019) wirkt sich diese Sorge negativer auf den Selbstwert aus, wenn die Ersetzung durch einen Menschen anstatt durch eine Maschine befürchtet wird. Entsprechend lässt sich vermuten, dass sich die Erkenntnisse zur präferierten Ersetzung durch eine Maschine anstatt durch einen Menschen auch auf menschenähnlich im Vergleich zu maschinell geframten Cobots übertragen lassen. Ferner impliziert dieses Ergebnis die Notwendigkeit, die gemeinhin vertretene Annahme einer generell vertrauenssteigernden Wirkung von Anthropomorphisierung einzuschränken und zu spezifizieren. Der statistisch signifikante Interaktionseffekt unterstreicht die Relevanz relationaler Aspekte in industriellen MRI-Settings und liefert eine hypothetische Erklärung für kontroverse experimentelle Befunde zu den Auswirkungen von Anthropomorphismus auf Vertrauen.

Das weitere Ziel, die realen Auswirkungen von Vertrauen sowie die gefundenen Framing-Effekte näher zu beleuchten und insbesondere die Übertragbarkeit auf andere Populationen zu überprüfen, erforderte die Durchführung einer weiteren Experimentalstudie. Diese zeichnet sich insbesondere durch die Integration eines anwendungsnahen, realistischen

Mensch-Cobot-Interaktionsszenarios innerhalb eines Präsenz-Settings aus. Die Stichprobe für dieses Experiment bestand hauptsächlich aus Studierenden mit tendenziell starkem Interesse an Robotern. Zunächst zeigte sich entgegen der Erwartung gemäß der ersten Experimentalstudie, dass die Lektüre der unterschiedlichen Framing-Texte keine Veränderung im Grad der wahrgenommenen Menschenähnlichkeit hervorrief. Folglich unterschied sich das Vertrauensniveau der beiden Experimentalgruppen nach dem Framing in der zweiten Experimentalstudie nicht signifikant voneinander. Vermutete Auswirkungen eines unterschiedlich starken Mensch-Cobot-Vertrauens auf die Wahrnehmung der realen Interaktion ließen sich nur durch eine nachträgliche Aufspaltung der Stichprobe untersuchen (FF4). Die deskriptive Betrachtung lieferte Hinweise darauf, dass ein höheres Vertrauensniveau mit einer positiveren Wahrnehmung der Interaktion einhergeht und sich auch im Verhalten widerspiegelt. Nachweisbare Korrelationen zwischen Vertrauenswerten und vereinzelt, verhaltensbasierten Items aus Technologieakzeptanzmodellen stützen diese Vermutung. Gleichwohl ist diese Evidenz als schwach zu bewerten, da die Ergebnisse nicht auf einer aktiven experimentellen Manipulation des Vertrauensniveaus beruhen. Die mangelnde Wirksamkeit des Framings ist auch der Grund dafür, dass keine empirischen Erkenntnisse zur Persistenz von Framing-Effekten über die Dauer einer realen MRI hinweg gewonnen werden konnten (FF5). Die stark unterschiedlichen Auswirkungen des Framings in beiden Experimentalstudien akzentuieren hingegen die Relevanz populationspezifischer Merkmale (FF6). Eine vergleichende Analyse der teilnehmenden Produktionsmitarbeitenden und Studierenden offenbarte signifikante Unterschiede in beinahe allen relevanten untersuchten Merkmalen wie der Technikkompetenzüberzeugung, der Vorkenntnisse in Bezug auf Robotik, des initialen Vertrauensniveaus und der Grundeinstellung gegenüber Robotern. Auf Basis bestehender theoretischer Überlegungen liegt die Vermutung nahe, dass Framing-Effekte besonders bei weniger technisch versierten Personen eine hohe Wirkkraft entfalten. Dennoch ist auch die Differenz zwischen dem Online-Setting aus der ersten Experimentalstudie, das besonders kognitive Vertrauenskomponenten adressiert, und dem Präsenz-Setting aus der zweiten Experimentalstudie, das verstärkt affektive Komponenten anspricht, als mögliche Erklärung einzubeziehen. Vor dem Hintergrund dieser Ergebnisse lässt sich die externe Validität der überwiegenden Anzahl bisheriger Studien problematisieren, die auf Studierende als Stichprobe zurückgreifen und auf reale Interaktionsszenarien verzichten.

Diese experimentell gewonnenen Erkenntnisse mündeten in ein umfassendes Verständnis der relevanten theoretischen Konstrukte der kognitiven MRI und in die Vertiefung sowie Integration bestehender Theorieansätze. Um der Schnelllebigkeit des Forschungsfeldes Rechnung zu tragen, wurden die Erkenntnisse dieser Dissertation in die aktuelle Forschung eingebettet. Das Mensch-Cobot-Vertrauen am Arbeitsplatz wird dabei als ein Prozess der Auflösung multipler Ambiguitäten betrachtet, die auf verschiedenen Ebenen durch die Ausgestaltung der Cobots, deren Einsatzkontext und die verbreiteten Narrative entstehen. Das Vertrauen als Festlegung auf eine Zukunftserwartung überwindet die kognitive

Komplexität, die sich aus der wahrgenommenen Ambiguität ergibt. Die Vertrauensbildung ist dabei ein kontinuierlicher Prozess, der typischerweise verschiedene Phasen der Ausbildung, der Vulnerabilität und der Stabilisierung durchläuft. In diesen Phasen verändern unterschiedliche Einflussfaktoren ihr Gewicht. Während das initiale Vertrauen stark von individuellen Faktoren abhängt und eine hohe Varianz aufweist, konvergiert das Vertrauen tendenziell entweder durch gezielte Informationsvermittlung oder im Verlauf der Interaktion durch Akkumulation konsistenter Erfahrungswerte. Gängige Erhebungsmethoden auf Basis von Selbsteinschätzungen vermögen diese Kontinuität kaum abzubilden und erschweren damit die zunehmend in den Fokus rückende Untersuchung von dynamischen Vertrauensbeschädigungs- und -reparaturprozessen, bspw. im Falle von unerwarteten Fehlfunktionen des Roboters. Ferner offenbart sich zunehmend der multidimensionale Charakter des Vertrauenskonstrukts. Dieser drückt sich bspw. in Differenzierungen in affektive und kognitive oder leistungs- und relationsorientierte Komponenten aus, deren Relevanz kontextspezifisch unterschiedlich zu beurteilen ist. Gerade im verpflichtenden Nutzungskontext am industriellen Arbeitsplatz, bei dem der Cobot oft nur stellvertretend für indirekt beteiligte Personen, Institutionen und Systeme als vertrauensnehmende Instanz auftritt, erfordert diese Multidimensionalität eine differenzierte Betrachtung. Das gilt insbesondere, da die Reduktion auf leistungsorientiertes Vertrauen verkennt, dass bspw. Sorgen vor sozialer Isolation, relationale Verortungen und Selbstdeutungen sowie implizite Vermenschlichungsprozesse am industriellen Arbeitsplatz eine gewichtige Rolle spielen. Auch die Betrachtung von Vertrauen und Misstrauen als separate und mit unterschiedlichen Faktoren assoziierte Konstrukte erscheint für empirische Erhebungen gewinnbringend. Ferner stellt die Möglichkeit zur realen Interaktion mit einem Cobot einen wichtigen vertrauensstiftenden Faktor dar, der sich mit dem sozialpsychologischen Effekt der Darbietungshäufigkeit in Verbindung bringen lässt. Darüber hinaus bewirkt die imaginierte Interaktion mit einem Cobot, die bei der Lektüre relevanter Informationen durch Personen ohne vorherige Erfahrungswerte entstehen kann, bereits eine relevante, wenngleich weniger starke Vertrauenssteigerung.

Die bis dato kaum empirisch erhobene Mensch-Roboter-Relation nimmt eine zentrale Rolle bei der Betrachtung von MRI-Situationen am Arbeitsplatz ein. Die Relevanz ergibt sich dabei primär aus der Brisanz der virulenten Angst vor Arbeitsplatzverlust und der damit verbundenen unterschiedlichen psychischen Reaktions- und Bewältigungsmuster. Nachdem bisher vorwiegend Fragen nach dem ontologischen Status von Robotern untersucht wurden, rücken allmählich phänomenologisch-relationale Fragen in den Fokus. Während in der Roboterethik bereits von einer relationalen Wende gesprochen wird, sind diese Fragen empirisch allerdings noch wenig untersucht. Dahinter steht die Annahme, dass für das Gelingen einer MRI und für deren subjektive Wahrnehmung die Art und Weise relevant ist, wie sich Mitarbeitende mental zu dem beteiligten Roboter in Beziehung setzen. Die relationalen Betrachtungen gehen dabei über die verkürzende Dichotomie zwischen dem Roboter als Werkzeug oder als Teamkolleg:in hinaus. Neuere Forschungsarbeiten

betonen zwar, dass Roboter inzwischen vermehrt als Teammitglieder auftreten. Damit ist die qualitative Auffassung der Beziehung allerdings kaum spezifiziert. Im Kontext mit Robotertechnologie am Arbeitsplatz stellt das Spannungsverhältnis zwischen kooperations- und konkurrenzorientierten Beziehungskonzeptionen einen entscheidenden Faktor dar.

Überdies moderiert die Mensch-Roboter-Relation die Auswirkungen von Vermenschlichung auf Vertrauen. Dieser Umstand betont die Kontextabhängigkeit von Anthropomorphismus-Effekten. Vermenschlichende Zuschreibungen geschehen am Arbeitsplatz in erster Linie implizit. Das unterstreicht die Notwendigkeit des Einsatzes indirekter Messmethoden, die über verbreitete Selbstauskunftsskalen hinausgehen. Ferner sind die Auswirkungen von Vermenschlichung ebenfalls multidimensional zu betrachten und wirken sich vermutlich unterschiedlich auf verschiedene Vertrauensdimensionen aus. Dabei ist die unterschiedliche Aktivierung des Schemas der perfekten Automation zu beachten, die Implikationen für die spätere Fehlertoleranz mit sich bringt. Eine etwaige initial vertrauensstiftende Wirkung von Vermenschlichung ist oftmals mit verschiedenartigen Erwartungshaltungen verknüpft, die nur in gewissen Interaktionskontexten wünschenswert sind und sich im Verlauf einer realen Interaktion schnell als irrig erweisen und ursprüngliche positive Effekte ins Gegenteil verkehren können.

Das sprachliche Framing wirkte bei den Produktionsmitarbeitenden in bemerkenswerter Weise auf die wahrgenommene Menschenähnlichkeit des Cobots und die Kooperationsorientierung der Mensch-Cobot-Relation. Zur Erklärung dieser starken Auswirkungen subtiler Variationen lieferte u. a. das Modell der Frame-Selektion aus der Soziologie relevante Ansatzpunkte. Auch in Hinblick auf die Wirksamkeit des Framings scheint die wahrgenommene Ambiguität des Cobots eine zentrale Rolle zu spielen. In Abwesenheit eindeutiger diskriminierender Hinweisreize fungieren die sprachlichen Merkmale als Zünglein an der Waage, um in einem automatisch-spontanen Verarbeitungsmodus ohne vertiefte, rationale und vollständige Elaboration der vorgefundenen Situation ein mentales Modell zu aktivieren und anzuwenden. Insofern determinieren die Vorerfahrungen, die wahrgenommene Ambiguität, die individuelle Ambiguitätsaversion und die Anreize zu einer reflektierten und elaborierten Informationsverarbeitung die Wirkung des Framings. Entsprechend verdeutlichte die zweite Experimentalstudie, dass die Wirkkraft des Framings personenindividuell sehr unterschiedlich ausfällt. Insbesondere Personen mit erheblichen Vorkenntnissen, einem gefestigten Bild von Robotern, einer hohen Ambiguitätstoleranz oder einem starken Bedürfnis nach elaborierter Verarbeitung lassen sich mutmaßlich weniger stark durch sprachliches Framing beeinflussen.

Aus diesen inhaltlichen Erkenntnissen ergaben sich praktische Implikationen, die einerseits aus unternehmerischer und andererseits aus gesellschaftlicher Sicht dargelegt wurden. Für Unternehmen, die eine Cobot-Einführung planen, sind u. a. personenspezifisch angepasste Schulungs- und Kommunikationsmaßnahmen inklusive praktischer Erprobungsräume und

einer sensiblen Sprachverwendung ratsam, um eine angemessene Vertrauenskalibrierung zu erzielen. Aus gesellschaftlicher Sicht sind insbesondere die Erkenntnisse zur Wirkkraft des sprachlichen Framings relevant. Der Verlauf öffentlicher Debatten und die positive Darstellung moderner Technologien bergen die Gefahr, dass sich unrealistische Erwartungen und ein kollektives Unterlegenheitsgefühl manifestieren. Diese Auswirkungen sind insbesondere in Hinblick auf die wachsende Präsenz von Robotern in verschiedenen Alltagsumgebungen kritisch zu sehen und könnten durch Ansätze des ehrlichen Designs, durch eine reflektierte achtsame Sprachverwendung oder durch umfassende Bildungsmaßnahmen vermieden bzw. abgemildert werden.

Grundsätzlich lassen sich die Ergebnisse dieser Dissertation auf andere Arten von interaktiven Robotern und vergleichbare moderne Technologien übertragen. Insbesondere die Wahrnehmung von KI-Anwendungen lässt sich aufgrund der intransparenten Funktionsweise sowie der Immaterialität dieser Anwendungen und dem gebräuchlichen anthropomorphen Vokabular in der öffentlichen Debatte mutmaßlich besonders stark durch sprachliches Framing beeinflussen. Bei der Übertragung der Ergebnisse auf andere Personengruppen ist die experimentell angezeigte individuell unterschiedliche und auch durch die entsprechenden Kulturkreise geprägte Grundeinstellung gegenüber Robotern zu berücksichtigen. Auch die Generalisierung auf andere Nutzungskontexte wie bspw. die Service-Robotik steht unter gewissen Vorbehalten. Insbesondere ist zu beachten, dass es sich dabei im Regelfall um keinen verpflichtenden Nutzungskontext handelt, sondern die Nutzung nur zustande kommt, sofern ein ausreichend hohes initiales Vertrauen besteht. Ferner ist die Relation nicht von der Ambivalenz zwischen Unterstützung und Ersetzung gekennzeichnet und somit um eine Komplexitätsdimension reduziert.

Im Rahmen einer methodischen Reflexion wurde das gewählte methodenpluralistische, realitätsnahe, interdisziplinäre Vorgehen unter Verwendung validierter Selbstauskunftsskalen und indirekter Messmethoden als zweckdienlich bewertet. Viele übliche Kritikpunkte an bisheriger MRI-Forschung ließen sich damit umgehen. Nichtsdestotrotz bestehen methodische Einschränkungen. Diese betreffen die begrenzte externe Validität durch das Online-Setting in der ersten sowie durch den Rückgriff auf Studierende in der zweiten Experimentalstudie, die begrenzte Repräsentativität der Stichprobe, die Messung von Vertrauen und Misstrauen als singuläre Konstrukte, die ausschließliche Messung der Interaktionswahrnehmung über Selbstauskunftsskalen, die Operationalisierung der Mensch-Cobot-Relation über eine Ad-hoc-Skala sowie die Möglichkeit unterschiedlicher Informationsgehalte bei den Framing-Texten. Aus diesen methodischen Einschränkungen und dem begrenzten Anwendungsbereich der Ergebnisse ergeben sich Anknüpfungspunkte für weitere Forschungsarbeiten. In erster Linie ist in diesem Kontext die differenzierte Erforschung unterschiedlicher Vertrauensdimensionen im zeitdynamischen Veränderungsprozess und unter Einbezug einer differenziert erfassten Mensch-Cobot-Relation zu nennen. Ferner sollten die kontext- und personenspezifischen Wirkbedingungen von Framing näher

ergründet werden. Eine Replikation der durchgeführten Experimentalstudien könnte deren Repräsentativität erhöhen und eine bessere Vergleichbarkeit erzielen, um u. a. rein methodische Erklärungen für das Ausbleiben der Framing-Effekte in der zweiten Experimentalstudie auszuschließen. Ein übergeordnetes Ziel stellt eine stärkere Integration bestehender Theorieansätze unter Berücksichtigung kontextspezifischer Einflüsse und der Wechselbeziehungen verschiedener Faktoren dar.

Als Fazit lässt sich konstatieren, dass das Mensch-Cobot-Vertrauen am Arbeitsplatz eine entscheidende Rolle für den Erfolg betrieblicher Cobot-Einführungen spielt und besonders angesichts hybrider Roboter und der Spezifika des Arbeitsplatzkontexts vielfältige Dimensionen, Einflussfaktoren und Auswirkungen umfasst. Die hohe Relevanz des sprachlichen Framings sowie phänomenologisch-relationaler Verortungen der Mitarbeitenden wird in diesem funktionalen Kontext häufig unterschätzt und erfordert eine tiefergehende Untersuchung. Mensch-Roboter-Relationen am Arbeitsplatz und damit zusammenhängende soziale Zuschreibungen scheinen komplexer zu sein als traditionelle und verkürzende Darstellungen nahelegen und können damit eine wichtige Schablone für Mensch-Roboter-Beziehungen in anderen Anwendungs- und Lebensbereichen repräsentieren. Die Forschung befindet sich noch in einer Phase, in der sie die Multidimensionalität, die Kontextabhängigkeit und die damit verbundene Komplexität des Mensch-Roboter-Vertrauens allmählich offenlegt. Inzwischen gestaltet sich die Forschung zunehmend ganzheitlicher, da menschenzentrierte und kontextuelle Faktoren sowie Top-Down-Prozesse stärker einbezogen werden, wenngleich diese weiterhin unterrepräsentiert zu sein scheinen. Die Anwendung von Theorien aus verschiedenen Forschungsdisziplinen wie der Psychologie, Soziologie und Philosophie erwies sich als gewinnbringend, was die Notwendigkeit nach interdisziplinärer Forschung unterstreicht. Die methodische Reflexion betont die bestehenden Forderungen nach stärkeren experimentellen Designs, um einerseits zu belastbaren und andererseits zu in die Realität übertragbaren Resultaten zu gelangen. Ein Prozess der Synthese bisheriger Forschungsergebnisse zur Entwicklung eines zusammenhängenden Theoriebildes scheint zunehmend angezeigt und kann an die in dieser Dissertation skizzierten produktiven Denkfiguren und Theorieansätze anschließen. Bisher existieren noch viele konfligierende Forschungsergebnisse, die einerseits stark unterschiedlichen experimentellen Designs und andererseits der hohen Komplexität und Kontextabhängigkeit der entsprechenden Phänomene geschuldet sind. Insofern besteht der Mehrwert dieser Dissertation neben der Beantwortung der Forschungsfragen darin, bisher zu wenig beachtete Perspektiven, Forschungsbedarfe und methodische Anforderungen zu identifizieren und produktive Anknüpfungspunkte für Theorieansätze aus verschiedenen Fachdisziplinen aufzuzeigen, die eine interdisziplinäre Herangehensweise nahelegen.

Literaturverzeichnis

- Aaltonen, I. & Salmi, T. (2019). Experiences and expectations of collaborative robots in industry and academia: barriers and development needs. *Procedia Manufacturing*, 38, 1151–1158. <https://doi.org/10.1016/j.promfg.2020.01.204>
- Aaltonen, I., Salmi, T. & Marstio, I. (2018). Refining levels of collaboration to support the design and evaluation of human-robot interaction in the manufacturing industry. *Procedia CIRP*, 72, 93–98. <https://doi.org/10.1016/j.procir.2018.03.214>
- Abelson, R. P. (1981). Psychological status of the script concept. *American Psychologist*, 36(7), 715–729. <https://doi.org/10.1037/0003-066X.36.7.715>
- The ABOT Database. The Anthropomorphic Robot Database.* (2019). Zugriff am 19.02.2021. Verfügbar unter: <http://abotdatabase.info/collection>
- Abubshait, A., Momen, A. & Wiese, E. (2017). Seeing human: Do individual differences modulate the Uncanny Valley? *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 61(1), 870–874. <https://doi.org/10.1177/1541931213601690>
- Abubshait, A. & Wiese, E. (2017). You Look Human, But Act Like a Machine: Agent Appearance and Behavior Modulate Different Aspects of Human-Robot Interaction. *Frontiers in Psychology*, 8, Aufsatz 1393. <https://doi.org/10.3389/fpsyg.2017.01393>
- Ahn, S. J., Fox, J., Dale, K. R. & Avant, J. A. (2015). Framing Virtual Experiences. *Communication Research*, 42(6), 839–863. <https://doi.org/10.1177/0093650214534973>
- AI Now. (2016). *The AI Now Report 2016. The Social and Economic Implications of Artificial Intelligence Technologies in the Near-Term.* Zugriff am 05.09.2017. Verfügbar unter: https://artificialintelligenzenow.com/media/documents/AINowSummaryReport_3_RpmwKHu.pdf
- Airenti, G. (2015). The Cognitive Bases of Anthropomorphism. From Relatedness to Empathy. *International Journal of Social Robotics*, 7(1), 117–127. <https://doi.org/10.1007/s12369-014-0263-x>
- Ajzen, I. (1985). *From Intentions to Actions: A Theory of Planned Behavior.* Berlin/Heidelberg: Springer. https://doi.org/10.1007/978-3-642-69746-3_2
- Ajzen, I. & Fishbein, M. (1980). *Understanding attitudes and predicting social behavior.* Englewood Cliffs, NJ: Prentice-Hall.
- Akella, P., Peshkin, M., Colgate, E., Wannasuphprasit, W., Nagesh, N., Wells, J. et al. (1999, Mai). Cobots for the automobile assembly line. In *Proceedings of the 1999 IEEE International Conference on Robotics & Automation (May 10-15 1999)* (S. 728–733). Piscataway.
- Alarcon, G. M., Capiola, A. & Pfahler, M. D. (2021). The role of human personality on trust in human-robot interaction. In C. S. Nam & J. B. Lyons (Eds.), *Trust in Human-*

- Robot Interaction* (S. 159–178). Elsevier Academic Press.
<https://doi.org/10.1016/B978-0-12-819472-0.00007-1>
- 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, 274–286. <https://doi.org/10.1016/j.chb.2019.07.031>
- Atkinson, D., Hancock, P. A., Hoffman, R. R., Lee, J. D., Rovira, E., Stokes, C. et al. (2012). Trust in Computers and Robots: The Uses and Boundaries of the Analogy to Interpersonal Trust. In *Proceedings of the Human Factors and Ergonomics Society 56th Annual Meeting* (S. 303–307). <https://doi.org/10.1177/1071181312561071>
- Baden, C. & Lecheler, S. (2012). Fleeting, Fading, or Far-Reaching? A Knowledge-Based Model of the Persistence of Framing Effects. *Communication Theory*, 22(4), 359–382. <https://doi.org/10.1111/j.1468-2885.2012.01413.x>
- Baker, A. L., Phillips, E. K., Ullman, D. & Keebler, J. R. (2018). Toward an Understanding of Trust Repair in Human-Robot Interaction. *ACM Transactions on Interactive Intelligent Systems*, 8(4), 1–30. <https://doi.org/10.1145/3181671>
- Baker, L. J., Hymel, A. M. & Levin, D. T. (2018). Anthropomorphism and Intentionality Improve Memory for Events. *Discourse Processes*, 55(3), 241–255. <https://doi.org/10.1080/0163853X.2016.1223517>
- Bargh, J. A., Chen, M. & Burrows, L. (1996). Automaticity of social behavior: Direct effects of trait construct and stereotype activation on action. *Journal of Personality and Social Psychology*, 71(2), 230–244. <https://doi.org/10.1037/0022-3514.71.2.230>
- Bartlett, C. F. (1932). *Remembering*. London: Cambridge University Press.
- Bartneck, C. (2019). *Negative Attitudes Towards Robots Scale (NARS) Translations*. Zugriff am 28.02.2020. Verfügbar unter: <http://www.bartneck.de/2019/03/11/negative-attitudes-towards-robots-scale-nars-translations/>
- Bartneck, C. & Forlizzi, J. (2004). A design-centred framework for social human-robot interaction. In *RO-MAN 2004. 13th IEEE International Workshop on Robot and Human Interactive Communication: proceedings: September 20-22, 2004, Kurashiki, Okayama Japan at Kurashiki Ivy Square* (S. 591–594). Piscataway, N.J: IEEE.
- Bartneck, C., Kanda, T., Mubin, O. & Al Mahmud, A. (2009). Does the Design of a Robot Influence Its Animacy and Perceived Intelligence? *International Journal of Social Robotics*, 1(2), 195–204. <https://doi.org/10.1007/s12369-009-0013-7>
- Bartneck, C., Kulić, D., Croft, E. & Zoghbi, S. (2009). Measurement Instruments for the Anthropomorphism, Animacy, Likeability, Perceived Intelligence, and Perceived Safety of Robots. *International Journal of Social Robotics*, 1(1), 71–81. <https://doi.org/10.1007/s12369-008-0001-3>
- Bartneck, C., Rosalia, C., Menges, R. & Deckers, I. (2005). Robot Abuse – A Limitation of the Media Equation. *Interact 2005 Workshop on Abuse, Rome*.

- Bartneck, C., Suzuki, T., Kanda, T. & Nomura, T. (2006). The influence of people's culture and prior experiences with Aibo on their attitude towards robots. *AI & SOCIETY*, 21(1-2), 217–230. <https://doi.org/10.1007/s00146-006-0052-7>
- Bauer, A., Wollherr, D. & Buss, M. (2007). Human-Robot Collaboration: A Survey. *International Journal of Humanoid Robotics*, 05(01), 47–66. <https://doi.org/10.1142/S0219843608001303>
- Bauer, M., Knape, J., Koch, P. & Winkler, S. (2010). Dimensionen der Ambiguität. *Zeitschrift für Literaturwissenschaft und Linguistik*, 40(158), 7–75.
- Baumgartner, M. (2020). *Vertrauensfördernde Maßnahmen für die industrielle Mensch-Roboter-Interaktion. Eine Analyse der sprachlichen Beeinflussbarkeit des Vertrauens von Mitarbeitern in Roboter*. Master-Thesis. Hochschule Karlsruhe, Karlsruhe.
- Baumgartner, M., Kopp, T. & Kinkel, S. (2020). Industrielle Mensch-Roboter-Kollaboration in KMU. KMU unterschätzen das Potenzial der Mensch-Roboter-Interaktion. *wt Werkstattstechnik online*, 110(3), 146–150.
- Becker, H., Scheermesser, M., Früh, M., Treusch, Y., Auerbach, H., Hüppi, R. A. et al. (2013). *Robotik in Betreuung und Gesundheitsversorgung* (TA-SWISS, Bd. 58). Zürich: vdf. <https://doi.org/10.3218/3521-6>
- Beckert, B., Buschak, D., Graf, B., Hägele, M., Jäger, A., Moll, C. et al. (2016). *Automatisierung und Robotik-Systeme. Studien zum deutschen Innovationssystem Nr. 11-2016* (Expertenkommission Forschung und Innovation (EFI), Hrsg.). Karlsruhe.
- Beckmann, M., Mackenbrock, T., Pies, I. & Sardison, M. (2004). *Mentale Modelle und Vertrauensbildung. Eine wirtschaftsethische Analyse*. Diskussionspapier (2004-9).
- Bendel, O. (2019a). Die programmierte Moral. In C. Woopen & M. Jannes (Hrsg.), *Roboter in der Gesellschaft. Technische Möglichkeiten und menschliche Verantwortung* (Schriften zu Gesundheit und Gesellschaft - Studies on Health and Society, Bd. 2, 1. Auflage 2019, S. 35–52). Berlin: Springer.
- Bendel, O. (2019b). Sexroboter aus Sicht der Maschinenethik. In O. Bendel (Hrsg.), *Handbuch Maschinenethik* (Springer Reference Geisteswissenschaften, Bd. 6, S. 1–19). Wiesbaden: Springer Fachmedien Wiesbaden. https://doi.org/10.1007/978-3-658-17484-2_22-1
- Bender, M., Braun, M., Rally, P. & Scholtz, O. (2016). *Leichtbauroboter in der manuellen Montage - einfach einfach anfangen. Erste Erfahrungen von Anwenderunternehmen* (Bauer, W., Hrsg.). Fraunhofer IAQ.
- Bergamasco, M. & Herr, H. (2016). Human-Robot Augmentation. In B. Siciliano & O. Khatib (Eds.), *Springer handbook of robotics* (2nd ed., S. 1875–1906). Berlin: Springer.
- Bernotat, J., Eyssel, F. & Sachse, J. (2019). The (Fe)male Robot: How Robot Body Shape Impacts First Impressions and Trust Towards Robots. *International Journal of Social Robotics*, 85(4), 1–13. <https://doi.org/10.1007/s12369-019-00562-7>

- Bethel, C. L. & Murphy, R. R. (2010). Review of Human Studies Methods in HRI and Recommendations. *International Journal of Social Robotics*, 2(4), 347–359. <https://doi.org/10.1007/s12369-010-0064-9>
- Billard, A. & Dautenhahn, K. (1997). *Grounding communication in situated, social robots. Technical Report*. Zugriff am 19.03.2019. Verfügbar unter: <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.42.3233&rep=rep1&type=pdf>
- Billings, D. R., Schaefer, K. E., Chen, J. Y. C. & Hancock, P. A. (2012). Human-robot interaction: Developing trust in robots. In H. A. Yanco (ed.), *Proceedings of the seventh annual ACM/IEEE international conference on Human-Robot Interaction* (S. 109–110). New York, NY: ACM.
- Bilotta, E., Gabriele, L., Servidio, R. & Tavernise, A. (2007). Investigating mental representations in children interacting with small mobile robots. *Conference ICL2007 (September 26-28 2007, Villach, Österreich)*.
- Bloss, R. (2016). Collaborative robots are rapidly providing major improvements in productivity, safety, programing ease, portability and cost while addressing many new applications. *Industrial Robot: An International Journal*, 43(5), 481–494.
- Bøgh, S., Hvilshøj, M., Kristiansen, M. & Madsen, O. (2012). Identifying and evaluating suitable tasks for autonomous industrial mobile manipulators (AIMM). *The International Journal of Advanced Manufacturing Technology*, 61(5-8), 713–726. <https://doi.org/10.1007/s00170-011-3718-3>
- Bogner, A., Littig, B. & Menz, W. (2014). *Interviews mit Experten*. Wiesbaden: Springer Fachmedien Wiesbaden. <https://doi.org/10.1007/978-3-531-19416-5>
- Böhle, K. & Bopp, K. (2014). What a vision: The artificial companion. A piece of vision assessment including an expert survey. *Science, Technology and Innovation Studies*, 10(1), 155–186.
- Bonin, H., Gregory, T. & Zierahn, U. (2015). *Übertragung der Studie von Frey/Osborne: (2013) auf Deutschland. Forschungsbericht 455*. Berlin: Bundesministerium für Arbeit und Soziales.
- Borah, P. (2011). Conceptual Issues in Framing Theory: A Systematic Examination of a Decade's Literature. *Journal of Communication*, 61(2), 246–263. <https://doi.org/10.1111/j.1460-2466.2011.01539.x>
- Borenstein, J. (2011). Robots and the changing workforce. *AI & SOCIETY*, 26(1), 87–93. <https://doi.org/10.1007/s00146-009-0227-0>
- Bornstein, R. F. (1989). Exposure and affect: Overview and meta-analysis of research, 1968–1987. *Psychological Bulletin*, 106(2), 265–289. <https://doi.org/10.1037/0033-2909.106.2.265>
- Braitenberg, V. (1987). *Experiments in synthetic psychology* (Bradford books, 2nd pr). Cambridge, Mass.: The MIT Press.

- Brandstetter, J. (2017). *The Power of Robot Groups with a Focus on Persuasive and Linguistic Cues. Dissertation.* University of Canterbury, Christchurch, New Zealand. <https://doi.org/10.13140/RG.2.2.31707.26406>
- Brandstetter, J. & Bartneck, C. (2017). Robots will dominate the use of our language. *Adaptive Behavior*, 25(6), 1–14. <https://doi.org/10.1177/1059712317731606>
- Brandstetter, J., Beckner, C., Sandoval, E. B. & Bartneck, C. (2017). Persistent Lexical Entrainment in HRI. In B. Mutlu (ed.), *Proceedings of the 2017 ACM/IEEE International Conference on Human-Robot Interaction* (S. 63–72). New York, NY, USA: ACM Press.
- Breazeal, C. L. (2004). Social Interactions in HRI. The Robot View. *IEEE Transactions on Systems, Man and Cybernetics, Part C (Applications and Reviews)*, 34(2), 181–186. <https://doi.org/10.1109/TSMCC.2004.826268>
- Breazeal, C. L., Dautenhahn, K. & Kanda, T. (2016). Social Robotics. In B. Siciliano & O. Khatib (Eds.), *Springer handbook of robotics* (2nd ed., S. 1935–1972). Berlin: Springer.
- Breazeal, C. L., Kidd, C. D., Thomaz, A. L., Hoffman, G. & Berlin, M. (2005). Effects of nonverbal communication on efficiency and robustness in human-robot teamwork. In *IROS 2005. IEEE IRS/RSJ International Conference on Intelligent Robots and Systems (Augut 2-6 2005, Edmonton, Alberta, Kanada)* (S. 708–713). Piscataway, N.J.: IEEE.
- Broadbent, E. (2017). Interactions With Robots: The Truths We Reveal About Ourselves. *Annual Review of Psychology*, 68, 627–652. <https://doi.org/10.1146/annurev-psych-010416-043958>
- Broadbent, E., Lee, Y. I., Stafford, R. Q., Kuo, I. H. & MacDonald, B. A. (2011). Mental Schemas of Robots as More Human-Like Are Associated with Higher Blood Pressure and Negative Emotions in a Human-Robot Interaction. *International Journal of Social Robotics*, 3(3), 291–297. <https://doi.org/10.1007/s12369-011-0096-9>
- Bröhl, C., Nelles, J., Brandl, C., Mertens, A. & Nitsch, V. (2019). Human–Robot Collaboration Acceptance Model: Development and Comparison for Germany, Japan, China and the USA. *International Journal of Social Robotics*, 130(2), 42–59. <https://doi.org/10.1007/s12369-019-00593-0>
- Bröhl, C., Nelles, J., Brandl, C., Mertens, A. & Schlick, C. M. (2016). TAM Reloaded: A Technology Acceptance Model for Human-Robot Cooperation in Production Systems. In C. Stephanidis (Ed.), *HCI International 2016. Proceedings 18th international conference, HCI International 2016, Toronto, Canada, July 17-22, 2016* (Communications in computer and information science, vol. 617, S. 97–103). Switzerland: Springer. https://doi.org/10.1007/978-3-319-40548-3_16
- Bröhl, C., Nelles, J., Brandl, C., Mertens, A. & Schlick, C. M. (2017). Entwicklung und Analyse eines Akzeptanzmodells für die Mensch-Roboter-Kooperation in der

- Industrie. In *GfA-Frühjahrskongress 2017, Brugg und Zürich: Soziotechnische Gestaltung des digitalen Wandels*.
- Brooks, R. A. (2017). Die Verwechslung von Performanz und Kompetenz. In J. Brockman (Hrsg.), *Was sollen wir von künstlicher Intelligenz halten? Die führenden Wissenschaftler unserer Zeit über intelligente Maschinen* (S. 149–152). Frankfurt am Main: Fischer Taschenbuch.
- Bruno, G. & Antonelli, D. (2018). Dynamic task classification and assignment for the management of human-robot collaborative teams in workcells. *The International Journal of Advanced Manufacturing Technology*, 98(9), 2415–2427. <https://doi.org/10.1007/s00170-018-2400-4>
- Buchner, R., Wurhofer, D., Weiss, A. & Tscheligi, M. (2012, März). User experience of industrial robots over time. In *2012 7th ACM/IEEE International Conference on Human-Robot Interaction (HRI)* (S. 115–116).
- Burleigh, T. J. & Schoenherr, J. R. (2015). A reappraisal of the uncanny valley: categorical perception or frequency-based sensitization? *Frontiers in Psychology*, 5, Aufsatz 1488. <https://doi.org/10.3389/fpsyg.2014.01488>
- Cacciatore, M. A., Scheufele, D. A. & Iyengar, S. (2016). The End of Framing as we Know it ... and the Future of Media Effects. *Mass Communication and Society*, 19(1), 7–23. <https://doi.org/10.1080/15205436.2015.1068811>
- Cacioppo, J. T., Petty, R. E., Feinstein, J. A. & Jarvis, W. B. G. (1996). Dispositional differences in cognitive motivation: The life and times of individuals varying in need for cognition. *Psychological Bulletin*, 119(2), 197–253. <https://doi.org/10.1037/0033-2909.119.2.197>
- Caglioti, G. (1992). *The Dynamics of Ambiguity*. Berlin/Heidelberg: Springer.
- Cameron, D., Aitken, J. M., Collins, E. C., Boorman, L., Chua, A., Fernando, S. et al. (2015). Framing Factors: The Importance of Context and the Individual in Understanding Trust in Human-Robot Interaction. In *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS) 2015*.
- Cameron, D., Loh, E. J., Chua, A., Collins, E. C., Aitken, J. M. & Law, J. (2016). Robot-stated limitations but not intentions promote user assistance. In *5th International Symposium on New Frontiers in Human-Robot Interaction* (S. 1–6).
- Campa, R. (2016). The Rise of Social Robots: A Review of the Recent Literature. *Journal of Evolution and Technology*, 26(1), 106–113.
- Campolo, A., Sanfilippo, M., Whittaker, M. & Crawford, K. 2017. *AI Now 2017 Report* (Selbst, A. & Barocas, S., eds.) (85). AI Now. <https://doi.org/10.14714/CP85.1413>
- Canellas, M. C., Miller, M. J., Razin, Y. S., Minotra, D., Bhattacharyya, R. & Haga, R. A. (2017). Framing Human-Automation Regulation: A new Modus Operandi from Cognitive Engineering. In *WeRobot 2017*.
- Čapek, K. (1920/2012). *R.U.R. (Rossum's universal robots)*. Maryland: Wildside Press.

- Cave, S., Coughlan, K. & Dihal, K. (2019). "Scary Robots". Examining Public Responses to AI. In V. Conitzer (ed.), *Proceedings of the 2019 AAAI/ACM Conference on AI, Ethics, and Society* (ACM Digital Library, S. 331–337). New York, USA: Association for Computing Machinery.
- Charalambous, G., Fletcher, S. R. & Webb, P. (2013). Human-Automation Collaboration in Manufacturing: Identifying Key Implementation Factors. In M. Anderson (ed.), *Contemporary ergonomics and human factors 2013. Proceedings of the international conference on Ergonomics & Human Factors 2013* (S. 301–306). Taylor & Francis.
- Charalambous, G., Fletcher, S. R. & Webb, P. (2016a). Development of a Human Factors Roadmap for the Successful Implementation of Industrial Human-Robot Collaboration. In C. M. Schlick & S. Trzciliński (Eds.), *Advances in Ergonomics of Manufacturing: Managing the Enterprise of the Future. Proceedings of the AHFE 2016 International Conference on Human Aspects of Advanced Manufacturing, July 27-31, 2016, Walt Disney World®, Florida, USA* (Advances in Intelligent Systems and Computing, vol. 490, S. 195–206). Cham: Springer International Publishing. https://doi.org/10.1007/978-3-319-41697-7_18
- Charalambous, G., Fletcher, S. R. & Webb, P. (2016b). The Development of a Scale to Evaluate Trust in Industrial Human-robot Collaboration. *International Journal of Social Robotics*, 8(2), 193–209. <https://doi.org/10.1007/s12369-015-0333-8>
- Charalambous, G., Fletcher, S. R. & Webb, P. (2017). The development of a Human Factors Readiness Level tool for implementing industrial human-robot collaboration. *The International Journal of Advanced Manufacturing Technology*, 91(5-8), 2465–2475. <https://doi.org/10.1007/s00170-016-9876-6>
- Charles, R. L., Charalambous, G. & Fletcher, S. R. (2015). Your New Colleague is a Robot. Is that okay? In S. Sharples, S. Shorrock & P. Waterson (Hrsg.), *Contemporary Ergonomics and Human Factors 2015. Proceedings of the International Conference on Ergonomics & Human Factors 2015* (S. 307–311). Taylor & Francis.
- Cherubini, A., Passama, R., Crosnier, A., Lasnier, A. & Fraisse, P. (2016). Collaborative manufacturing with physical human–robot interaction. *Robotics and Computer-Integrated Manufacturing*, 40, 1–13. <https://doi.org/10.1016/j.rcim.2015.12.007>
- Choi, H. & Swanson, N. (2021). Understanding worker trust in industrial robots for improving workplace safety. In C. S. Nam & J. B. Lyons (Eds.), *Trust in Human-Robot Interaction* (S. 123–141). Elsevier Academic Press. <https://doi.org/10.1016/B978-0-12-819472-0.00005-8>
- Choi, J. J. & Kwak, S. S. (2015). The Effect of Robot Appearance Types and Task Types on Service Evaluation of a Robot. In J. A. Adams (ed.), *Proceedings of the Tenth Annual ACM/IEEE International Conference on Human-Robot Interaction (Extended Abstracts)* (S. 223–224). New York, NY: ACM.
- Chong, D. & Druckman, J. N. (2007). Framing Theory. *Annual Review of Political Science*, 10(1), 103–126. <https://doi.org/10.1146/annurev.polisci.10.072805.103054>

- Claßen, K. (2013). *Zur Psychologie von Technikakzeptanz im höheren Lebensalter: Die Rolle von Technikgenerationen*. Dissertation. Ruprecht-Karls-Universität Heidelberg, Heidelberg. <https://doi.org/10.11588/heidok.00014295>
- Coeckelbergh, M. (2009). Personal Robots, Appearance, and Human Good. A Methodological Reflection on Roboethics. *International Journal of Social Robotics*, 1(3), 217–221. <https://doi.org/10.1007/s12369-009-0026-2>
- Coeckelbergh, M. (2010a). Robot rights? Towards a social-relational justification of moral consideration. *Ethics and Information Technology*, 12(3), 209–221. <https://doi.org/10.1007/s10676-010-9235-5>
- Coeckelbergh, M. (2010b). Talking to Robots: On the Linguistic Construction of Personal Human-Robot Relations. In M. H. Lamers & F. J. Verbeek (Eds.), *Human-Robot Personal Relationships. Third International Conference, HRPR 2010 (June 23-24 2010, Leiden, Niederlande), Revised Selected Papers* (vol. 38, S. 126–129).
- Coeckelbergh, M. (2011). You, robot. On the linguistic construction of artificial others. *AI & SOCIETY*, 26(1), 61–69. <https://doi.org/10.1007/s00146-010-0289-z>
- Coeckelbergh, M. (2012). *Growing moral relations. Critique of moral status ascription*. Basingstoke: Palgrave Macmillan. <https://doi.org/10.1057/9781137025968>
- Coeckelbergh, M. (2017). Language and technology: maps, bridges, and pathways. *AI & SOCIETY*, 32(2), 175–189. <https://doi.org/10.1007/s00146-015-0604-9>
- Coeckelbergh, M. (2018). Why Care About Robots? Empathy, Moral Standing, and the Language of Suffering. *Kairos. Journal of Philosophy & Science*, 20(1), 141–158. <https://doi.org/10.2478/kjps-2018-0007>
- Coeckelbergh, M. & Gunkel, D. J. (2014). Facing Animals: A Relational, Other-Oriented Approach to Moral Standing. *Journal of Agricultural and Environmental Ethics*, 27(5), 715–733. <https://doi.org/10.1007/s10806-013-9486-3>
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2. ed.). Hillsdale, NJ: Erlbaum.
- Cohn, A., Maréchal, M. A. & Noll, T. (2015). Bad Boys: How Criminal Identity Salience Affects Rule Violation. *The Review of Economic Studies*, 82(4), 1289–1308. <https://doi.org/10.1093/restud/rdv025>
- Compagna, D. & Muhl, C. (2012). Mensch-Roboter Interaktion – Status der technischen Entität, Kognitive (Des)Orientierung und Emergenzfunktion des Dritten. In *Muster und Verläufe der Mensch-Technik-Interaktivität. Band zum gleichnamigen Workshop am 17./18. Juni 2011 in Berlin* (S. 19–34) (Working Papers).
- Compagna, D., Weidemann, A., Marquardt, M. & Graf, P. (2016). Sociological and Biological Insights on How to Prevent the Reduction in Cognitive Activity that Stems from Robots Assuming Workloads in Human–Robot Cooperation. *Societies*, 6(4), 1–11. <https://doi.org/10.3390/soc6040029>
- Craenen, B., Deshmukh, A., Foster, M. E. & Vinciarelli, A. (2018). Do We Really Like Robots that Match our Personality? The Case of Big-Five Traits, Godspeed Scores

- and Robotic Gestures. In *27th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)* (S. 626–631). IEEE.
- Craik, K. J. W. (1967/1943). *The nature of explanation* (Repr. with postscript). Cambridge: Cambridge University Press.
- Damiano, L. & Dumouchel, P. (2018). Anthropomorphism in Human-Robot Co-evolution. *Frontiers in Psychology*, 9, Aufsatz 468. <https://doi.org/10.3389/fpsyg.2018.00468>
- Danaher, J. (2020). Welcoming Robots into the Moral Circle: A Defence of Ethical Behaviourism. *Science and Engineering Ethics*, 26(4), 2023–2049. <https://doi.org/10.1007/s11948-019-00119-x>
- Darling, K. (2016). Extending Legal Protection to Social Robots. The Effects of Anthropomorphism, Empathy, and Violent Behavior Towards Robotic Objects. In R. Calo, M. A. Froomkin & I. Kerr (Eds.), *Robot Law* (S. 213–232). Cheltenham, UK: Edward Elgar Publishing. <https://doi.org/10.2139/ssrn.2044797>
- Darling, K. (2017). "Who's Johnny?". Anthropomorphic Framing in Human-Robot Interaction, Integration, and Policy. In P. Lin, K. Abney & R. Jenkins (Eds.), *Robot ethics 2.0. From autonomous cars to artificial intelligence*. Oxford: Oxford University Press.
- Darling, K., Nandy, P. & Breazeal, C. L. (2015). Empathic concern and the effect of stories in human-robot interaction. In *2015 24th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN), Kobe, Japan, 2015* (S. 770–775). Piscataway, NJ: IEEE.
- Davis, F. D. (1985). *A technology acceptance model for empirically testing new end-user information systems: theory and results*. Dissertation. Massachusetts Institute of Technology, Sloan School of Management, Cambridge.
- De Graaf, M. M. A. (2016). An Ethical Evaluation of Human–Robot Relationships. *International Journal of Social Robotics*, 8(4), 589–598. <https://doi.org/10.1007/s12369-016-0368-5>
- De Graaf, M. M. A. & Ben Allouch, S. (2013). Exploring influencing variables for the acceptance of social robots. *Robotics and Autonomous Systems*, 61(12), 1476–1486. <https://doi.org/10.1016/j.robot.2013.07.007>
- De Graaf, M. M. A., Ben Allouch, S. & van Dijk, J. A. G. M. (2016). Long-term evaluation of a social robot in real homes. *Interaction Studies*, 17(3), 461–490. <https://doi.org/10.1075/is.17.3.08deg>
- De Visser, E. J., Monfort, S. S., McKendrick, R., Smith, M. A. B., McKnight, P. E., Krueger, F. et al. (2016). Almost human: Anthropomorphism increases trust resilience in cognitive agents. *Journal of Experimental Psychology. Applied*, 22(3), 331–349. <https://doi.org/10.1037/xap0000092>
- De Visser, E. J., Peeters, M. M. M., Jung, M. F., Kohn, S., Shaw, T. H., Pak, R. et al. (2020). Towards a Theory of Longitudinal Trust Calibration in Human–Robot Teams.

- International Journal of Social Robotics*, 12(2), Aufsatz 1393.
<https://doi.org/10.1007/s12369-019-00596-x>
- Decker, M. (2017). The next generation of robots for the next generation of humans? *Robotics and Autonomous Systems*, 88, 154–156. <https://doi.org/10.1016/j.robot.2016.11.005>
- Decker, M., Fischer, M. & Ott, I. (2017). Service Robotics and Human Labor: A first technology assessment of substitution and cooperation. *Robotics and Autonomous Systems*, 87, 348–354. <https://doi.org/10.1016/j.robot.2016.09.017>
- Dennett, D. C. (1971). Intentional Systems. *Journal of Philosophy*, 68(4), 87–106.
<https://doi.org/10.2307/2025382>
- Denzau, A. T. & North, D. C. (1994). Shared Mental Models: Ideologies and Institutions. *Kyklos*, 47(1), 3–31. <https://doi.org/10.1111/j.1467-6435.1994.tb02246.x>
- Desai, M., Kaniarasu, P., Medvedev, M., Steinfeld, A. & Yanco, H. A. (2013). Impact of robot failures and feedback on real-time trust. In *HRI 2013. Proceedings of the 8th ACM/IEEE International Conference on Human-Robot Interaction* (S. 251–258). Piscataway, NJ: IEEE.
- Devitt, K. S. (2018). Trustworthiness of Autonomous Systems. In H. A. Abbass, J. Scholz & D. J. Reid (Eds.), *Foundations of Trusted Autonomy* (vol. 117). Cham: Springer International Publishing.
- Dharshing, S., Hille, S. L. & Wüstenhagen, R. (2017). The Influence of Political Orientation on the Strength and Temporal Persistence of Policy Framing Effects. *Ecological Economics*, 142, 295–305. <https://doi.org/10.1016/j.ecolecon.2017.05.014>
- Dieber, B., Schlotzhauer, A. & Brandstötter, M. (2017). Safety & Security – Erfolgsfaktoren von sensitiven Robotertechnologien. *e & i*, 134(6), 299–303.
<https://doi.org/10.1007/s00502-017-0512-4>
- Dijksterhuis, A. & Bargh, J. A. (2001). The perception-behavior expressway: Automatic effects of social perception on social behavior. In *Advances in Experimental Social Psychology* (vol. 33, S. 1–40). Elsevier. [https://doi.org/10.1016/S0065-2601\(01\)80003-4](https://doi.org/10.1016/S0065-2601(01)80003-4)
- DIN EN ISO, 12100 (März 2011). *Sicherheit von Maschinen – Allgemeine Gestaltungsleitsätze – Risikobeurteilung und Risikominderung*. Berlin: Beuth Verlag GmbH.
- DIN EN ISO, 10218-1 (Januar 2012). *Industrieroboter – Sicherheitsanforderungen – Teil 1: Roboter*. Berlin: Beuth Verlag GmbH.
- DIN ISO/TR, 14121-2 (Februar 2013). *Sicherheit von Maschinen - Risikobeurteilung*. Berlin: Beuth Verlag GmbH.
- DIN EN ISO, 13849-1 (Juni 2016). *Sicherheit von Maschinen – Sicherheitsbezogene Teile von Steuerungen – Teil1: Allgemeine Gestaltungsleitsätze*. Berlin: Beuth Verlag GmbH.

- DIN EN ISO/ IEC, 27001 (Februar 2017). *Informationstechnik – Sicherheitsverfahren – Informationssicherheitsmanagementsysteme – Anforderungen*. Berlin: Beuth Verlag GmbH.
- DIN ISO/TS, 15066 (April 2017). *Roboter und Robotikgeräte - Kollaborierende Roboter*. Berlin: Beuth Verlag GmbH.
- DIN EN ISO 6385 (2016). *Grundsätze der Ergonomie für die Gestaltung von Arbeitssystemen*.
- Djuric, A. M., Urbanic, R. J. & Rickli, J. L. (2016). A Framework for Collaborative Robot (CoBot) Integration in Advanced Manufacturing Systems. *SAE International Journal of Materials and Manufacturing*, 9(2), 457–464.
<https://doi.org/10.4271/2016-01-0337>
- Döring, N. & Bortz, J. (2016). *Forschungsmethoden und Evaluation in den Sozial- und Humanwissenschaften* (Springer-Lehrbuch, 5. vollständig überarbeitete, aktualisierte und erweiterte Auflage). Berlin, Heidelberg: Springer. <https://doi.org/10.1007/978-3-642-41089-5>
- Dourish, P. (1996). Review of "The Media Equation". *Apple Labs Review*. Zugriff am 03.08.2020. Verfügbar unter: <https://www.dourish.com/publications/media-review.html>
- Duffy, B. R. (2003). Anthropomorphism and the social robot. *Robotics and Autonomous Systems*, 42(3-4), 177–190. [https://doi.org/10.1016/S0921-8890\(02\)00374-3](https://doi.org/10.1016/S0921-8890(02)00374-3)
- Dutke, S. (1994). *Mentale Modelle. Konstrukte des Wissens und Verstehens: kognitionspsychologische Grundlagen für die Software-Ergonomie* (Arbeit und Technik, Bd. 4). Göttingen: Verlag für Angewandte Psychologie.
- Dziak, J. J., Dierker, L. C. & Abar, B. (2020). The Interpretation of Statistical Power after the Data have been Gathered. *Current Psychology*, 39(3), 870–877.
<https://doi.org/10.1007/s12144-018-0018-1>
- Dzindolet, M. T., Pierce, L. G., Beck, H. P. & Dawe, L. A. (2002). The perceived utility of human and automated aids in a visual detection task. *Human Factors*, 44(1), 79–94. <https://doi.org/10.1518/0018720024494856>
- Edwards, A., Edwards, C., Westerman, D. & Spence, P. R. (2019). Initial expectations, interactions, and beyond with social robots. *Computers in Human Behavior*, 90, 308–314. <https://doi.org/10.1016/j.chb.2018.08.042>
- Edwards, C., Edwards, A., Spence, P. R. & Westerman, D. (2016). Initial Interaction Expectations with Robots: Testing the Human-To-Human Interaction Script. *Communication Studies*, 67(2), 227–238. <https://doi.org/10.1080/10510974.2015.1121899>
- Eklund, J. (2003). An extended framework for humans, technology and organization in interaction. In H. Luczak & K. J. Zink (Eds.), *Human factors in organizational design and management. Proceedings of the Seventh International Symposium on Human Factors in Organizational Design and Management, Aachen, Germany, October 1-2 2003* (S. 47–60). Santa Monica, Calif.: IEA Press; Mainz.

- El Makrini, I., Elprama, S. A., van den Bergh, J. & Lefeber, D. (2016). Design and Investigation of Social Robotic Coworkers in Factories. *International Journal of People-Oriented Programming*, 5(1), 19–38. <https://doi.org/10.4018/IJPOP.2016010102>
- El Makrini, I., Elprama, S. A., van den Bergh, J., Vanderborght, B., Knevels, A.-J., Jewell, C. I. et al. (2018). Working with Walt: How a Cobot Was Developed and Inserted on an Auto Assembly Line. *IEEE Robotics & Automation Magazine*, 25(2), 51–58. <https://doi.org/10.1109/MRA.2018.2815947>
- Ellsberg, D. (1961). Risk, Ambiguity, and the Savage Axioms. *The Quarterly Journal of Economics*, 75(4), 643–669. <https://doi.org/10.2307/1884324>
- Elprama, S. A., El Makrini, I., Vanderborght, B. & Jacobs, A. (2016). Acceptance of collaborative robots by factory workers: a pilot study on the importance of social cues of anthropomorphic robots. In *The 25th IEEE International Symposium on Robot and Human Interactive Communication* (S. 919–924). New York.
- Elprama, S. A., Jewell, C. I., Jacobs, A., El Makrini, I. & Vanderborght, B. (2017). Attitudes of Factory Workers towards Industrial and Collaborative Robots. In B. Mutlu, M. Tscheligi, A. Weiss & J. E. Young (eds.), *Proceedings of the Companion of the 2017 ACM/IEEE International Conference on Human-Robot Interaction - HRI '17* (S. 113–114). New York, NY, USA: ACM Press.
- Enquete-Kommission Künstliche Intelligenz. (2020). *Projektgruppe „KI und Arbeit, Bildung, Forschung“*. Zusammenfassung der vorläufigen Ergebnisse. Stand: 7. September 2020. Zugriff am 27.01.2021. Verfügbar unter: <https://www.bundestag.de/resource/blob/794582/e1f5a0d821f3fd33ac181ade9494e6c4/Pg-4-Projektgruppenbericht-data.pdf>
- Entman, R. M. (1993). Framing: Toward Clarification of a Fractured Paradigm. *Journal of Communication*, 43(4), 51–58. <https://doi.org/10.1111/j.1460-2466.1993.tb01304.x>
- Epley, N. (2018). A Mind like Mine: The Exceptionally Ordinary Underpinnings of Anthropomorphism. *Journal of the Association for Consumer Research*, 3(4), 591–598. <https://doi.org/10.1086/699516>
- Epley, N., Waytz, A., Akalis, S. & Cacioppo, J. T. (2008). When We Need A Human: Motivational Determinants of Anthropomorphism. *Social Cognition*, 26(2), 143–155. <https://doi.org/10.1521/soco.2008.26.2.143>
- Epley, N., Waytz, A. & Cacioppo, J. T. (2007). On seeing human: a three-factor theory of anthropomorphism. *Psychological Review*, 114(4), 864–886. <https://doi.org/10.1037/0033-295X.114.4.864>
- Erel, H., Hoffman, G. & Zuckerman, O. (2018). Interpreting non anthropomorphic robots social gestures. *HRI Workshop on Explainable Robotic 2018*. <https://doi.org/10.1145/1957656.1957668>
- Erickson, I., Robert, L. P., Crowston, K. & Nickerson, J. V. (2018). Workshop: Work in the Age of Intelligent Machines. In A. Forte, M. Prilla, A. Vivacqua, C. Müller & L. P. Robert (eds.), *GROUP 2018. Proceedings of the 2018 ACM International*

- Conference on Supporting Groupwork: January 7-10, 2018, Sanibel Island, FL, USA* (S. 359–361). New York: ACM.
- Esser, H. (2002). *Sinn und Kultur* (Soziologie, Bd. 6, Studienausgabe). Frankfurt/Main: Campus-Verlag.
- Esser, H. (2005). *Affektuelles Handeln: Emotionen und das Modell der Frame-Selektion* (Rationalitätskonzepte, Entscheidungsverhalten und ökonomische Modellierung 5).
- European Commission. (2015). *Autonomous Systems. Special Eurobarometer 427*. Zugriff am 10.08.2018.
- European Commission. (2017). *Attitudes towards the impact of digitisation and automation on daily life. Special Eurobarometer 460*. Zugriff am 05.01.2021.
- Evans, J. S. B. T. & Stanovich, K. E. (2013). Dual-Process Theories of Higher Cognition: Advancing the Debate. *Perspectives on Psychological Science*, 8(3), 223–241. <https://doi.org/10.1177/1745691612460685>
- Eyssel, F. & Hegel, F. (2012). (S)he's Got the Look: Gender Stereotyping of Robots. *Journal of Applied Social Psychology*, 42(9), 2213–2230. <https://doi.org/10.1111/j.1559-1816.2012.00937.x>
- Eyssel, F. & Kuchenbrandt, D. (2012). Social categorization of social robots. Anthropomorphism as a function of robot group membership. *The British Journal of Social Psychology*, 51(4), 724–731. <https://doi.org/10.1111/j.2044-8309.2011.02082.x>
- Eyssel, F. & Reich, N. (2013). Loneliness makes the heart grow fonder (of robots) — On the effects of loneliness on psychological anthropomorphism. In *HRI 2013. Proceedings of the 8th ACM/IEEE International Conference on Human-Robot Interaction* (S. 121–122). Piscataway, NJ: IEEE.
- Fehr, T. (2010). Big Five: Die fünf grundlegenden Dimensionen der Persönlichkeit und ihre 30 Facetten. In W. Simon (Hrsg.), *Persönlichkeitsmodelle und Persönlichkeits-tests. 15 Persönlichkeitsmodelle für Personalauswahl, Persönlichkeitsentwicklung, Training und Coaching* (GABAL professional training, 2. Aufl., S. 113–135). Offenbach: GABAL-Verl.
- Ferrari, F. & Eyssel, F. (2016). Toward a Hybrid Society. In A. Agah, A. M. Howard, M. A. Salichs, H. He & J.-J. Cabibihan (Eds.), *Social Robotics. Proceedings of the 8th International Conference, ICSR 2016 Kansas City, MO, USA, November 1–3, 2016* (S. 909–918). Cham: Springer International Publishing.
- Festinger, L. (1957). *A theory of cognitive dissonance* (1. publ). Stanford, Calif.: Stanford Univ. Press.
- Finch, H. (2005). Comparison of the Performance of Nonparametric and Parametric MANOVA Test Statistics when Assumptions Are Violated. *Methodology*, 1(1), 27–38. <https://doi.org/10.1027/1614-1881.1.1.27>
- Fine, G. A. & Holyfield, L. (1996). Secrecy, Trust, and Dangerous Leisure: Generating Group Cohesion in Voluntary Organizations. *Social Psychology Quarterly*, 59(1), 22. <https://doi.org/10.2307/2787117>

- Fink, J. (2012). Anthropomorphism and Human Likeness in the Design of Robots and Human-Robot Interaction. In S. S. Ge (Ed.), *Social robotics. Proceedings of the 4th International Conference, ICSR 2012, Chengdu, China, October 29-31, 2012* (vol. 7621, S. 199–208). Heidelberg: Springer. https://doi.org/10.1007/978-3-642-34103-8_20
- Fischer, K. (2011). Interpersonal Variation in Understanding Robots as Social Actors. In A. Billard (ed.), *Proceedings of the 6th international conference on Human-robot interaction* (S. 53–60). New York, NY: ACM.
- Fischer, K., Lohan, K. S. & Foth, K. (2012). Levels of embodiment: Linguistic Analyses of Factors Influencing HRI. In *Proceedings of the seventh annual ACM/IEEE international conference on Human-Robot Interaction* (S. 463–470). <https://doi.org/10.1145/2157689.2157839>
- Fishbein, M. & Ajzen, I. (1975). *Belief, attitude, intention and behavior. An introduction to theory and research*. Reading, Mass.: Addison-Wesley.
- Fitzpatrick, P., Harada, K., Kemp, C. C. & Matsumoto, Y. (2016). Humanoids. In B. Siciliano & O. Khatib (Eds.), *Springer handbook of robotics* (2nd ed., S. 1789–1818). Berlin: Springer.
- Fletcher, S. R. & Webb, P. (2017). Industrial Robot Ethics: The Challenges of Closer Human Collaboration in Future Manufacturing Systems. In M. I. Aldinhas Ferreira, J. Silva Sequeira, M. O. Tokhi, E. E. Kadar & G. S. Virk (Eds.), *A World with Robots. International Conference on Robot Ethics: ICRE 2015* (Intelligent Systems, Control and Automation, vol. 84, S. 159–169). Cham: Springer International Publishing. https://doi.org/10.1007/978-3-319-46667-5_12
- Flook, R., Shrinah, A., Wijnen, L., Eder, K., Melhuish, C. & Lemaignan, S. (2019). Social Cues in Robot Interaction, Trust and Acceptance. *Interaction Studies*, 20(3), 455–486. <https://doi.org/10.1075/is.18067.flo>
- Fogg, B. J. (2004). *Persuasive technology. Using computers to change what we think and do* (The Morgan Kaufmann series in interactive technologies). San Francisco: Morgan Kaufmann.
- Freelon, D. (2013). ReCal OIR: Ordinal, Interval, and Ratio Intercoder Reliability as a Web Service. *International Journal of Internet Science*, 8(1), 10–16.
- French, B., Duenser, A. & Heathcote, A. (2018). *Trust in Automation - A Literature Review report* (CSIRO Report EP184082). Australia: CSIRO.
- French, D. P., Sutton, S., Hennings, S. J., Mitchell, J., Wareham, N. J., Griffin, S. et al. (2005). The Importance of Affective Beliefs and Attitudes in the Theory of Planned Behavior: Predicting Intention to Increase Physical Activity. *Journal of Applied Social Psychology*, 35(9), 1824–1848. <https://doi.org/10.1111/j.1559-1816.2005.tb02197.x>

- Frennert, S. (2019). A Human Factor Approach to HRI. In M. A. Salichs, S. S. Ge, E. I. Barakova, J.-J. Cabibihan, A. R. Wagner, Á. Castro-González et al. (Eds.), *Social Robotics* (vol. 11876, S. 311–321). Cham: Springer International Publishing.
- Frennert, S., Efrting, H. & Östlund, B. (2017). Case Report: Implications of Doing Research on Socially Assistive Robots in Real Homes. *International Journal of Social Robotics*, 9(3), 401–415. <https://doi.org/10.1007/s12369-017-0396-9>
- Frey, C. B. & Osborne, M. A. (2013). *The Future of Employment: How susceptible are jobs to computerisation? Oxford Martin School (OMS) working paper*. Oxford. Zugriff am 15.03.2017. Verfügbar unter: https://www.oxfordmartin.ox.ac.uk/downloads/academic/The_Future_of_Employment.pdf
- Fussell, S. R., Kiesler, S., Setlock, L. D. & Yew, V. (2008). How people anthropomorphize robots. In *Proceedings of the Third ACM/IEEE Conference on Human-Robot Interaction, March 12 - 15, 2008, Amsterdam, The Netherlands; HRI 2008* (S. 145–152). New York, NY: ACM Press. <https://doi.org/10.1145/1349822.1349842>
- Gambino, A., Fox, J. & Ratan, R. (2020). Building a Stronger CASA: Extending the Computers Are Social Actors Paradigm. *Human-Machine Communication*, 1, 71–86. <https://doi.org/10.30658/hmc.1.5>
- Gaudiello, I., Zibetti, E., Lefort, S., Chetouani, M. & Ivaldi, S. (2016). Trust as indicator of robot functional and social acceptance. An experimental study on user conformation to iCub answers. *Computers in Human Behavior*, 61, 633–655. <https://doi.org/10.1016/j.chb.2016.03.057>
- Ghazizadeh, M., Lee, J. D. & Boyle, L. N. (2012). Extending the Technology Acceptance Model to assess automation. *Cognition, Technology & Work*, 14(1), 39–49. <https://doi.org/10.1007/s10111-011-0194-3>
- Giffi, C., Wellener, P., Dollar, B., Manolian, H. A., Monck, L. & Moutray, C. 2018. *Deloitte and The Manufacturing Institute skills gap and future of work study* (Deloitte Development LLC, ed.) (Deloitte Insights). London. Accessed 22.05.2020. Retrieved from <https://documents.deloitte.com/insights/2018DeloitteSkillsGapFoWManufacturing>
- Gläser, J. & Laudel, G. (2013). Life With and Without Coding: Two Methods for Early-Stage Data Analysis in Qualitative Research Aiming at Causal Explanations. *FQS - Forum Qualitative Sozialforschung*, 14(2), 1–37. <https://doi.org/10.17169/fqs-14.2.1886>
- Goetz, J., Kiesler, S. & Powers, A. (2012). Matching robot appearance and behavior to tasks to improve human-robot cooperation. In *The 12th IEEE International Workshop on Robot and Human Interactive Communication, ROMAN 2003. Proceedings*. (S. 55–60). <https://doi.org/10.1109/ROMAN.2003.1251796>
- Gombolay, M. C., Bair, A., Huang, C. & Shah, J. A. (2017). Computational design of mixed-initiative human–robot teaming that considers human factors. Situational

- awareness, workload, and workflow preferences. *The International Journal of Robotics Research*, 36(5-7), 597–617. <https://doi.org/10.1177/0278364916688255>
- Gompei, T. & Umemuro, H. (2018). Factors and Development of Cognitive and Affective Trust on Social Robots. In S. S. Ge (Ed.), *Social robotics. Proceedings of the 10th International Conference, ICSR 2018, Qingdao, China, November 28 - 30, 2018* (vol. 11357, S. 45–54). Cham: Springer. https://doi.org/10.1007/978-3-030-05204-1_5
- Görke, M., Blankemeyer, S., Pischke, D., Oubari, A., Raatz, A. & Nyhuis, P. (2017). Sichere und akzeptierte Kollaboration von Mensch und Maschine. *ZWF Zeitschrift für wirtschaftlichen Fabrikbetrieb*, 112(1-2), 41–45. <https://doi.org/10.3139/104.111668>
- Gottschling, V. (2006). Visual Imagery, Mental Models, and Reasoning. In C. Held, M. Knauff & G. Vosgerau (Eds.), *Mental models and the mind. Current developments in cognitive psychology, neuroscience, and philosophy of mind* (Advances in psychology, vol. 138, 1st ed., S. 211–235). Amsterdam: Elsevier.
- Grahn, S., Gopinath, V., Wang, X. V. & Johansen, K. (2018). Exploring a Model for Production System Design to Utilize Large Robots in Human-Robot Collaborative Assembly Cells. *Procedia Manufacturing*, 25, 612–619. <https://doi.org/10.1016/j.promfg.2018.06.094>
- Granulo, A., Fuchs, C. & Puntoni, S. (2019). Psychological reactions to human versus robotic job replacement. *Nature Human Behaviour*, 3(10), 1062–1069. <https://doi.org/10.1038/s41562-019-0670-y>
- Gray, H. M., Gray, K. & Wegner, D. M. (2007). Dimensions of mind perception. *Science*, 315(5812), 619. <https://doi.org/10.1126/science.1134475>
- Groom, V. & Nass, C. I. (2007). Can robots be teammates? Benchmarks in human-robot teams. *Interaction Studies*, 8(3), 483–500.
- Groom, V., Srinivasan, V., Bethel, C. L., Murphy, R. R., Dole, L. & Nass, C. I. (2011). Responses to robot social roles and social role framing. *2011 International Conference on Collaboration Technologies and Systems (CTS), Philadelphia, PA, USA*, 194–203. <https://doi.org/10.1109/CTS.2011.5928687>
- Groom, V., Takayama, L., Ochi, P. & Nass, C. I. (2009). I Am My Robot: The Impact of Robot-building and Robot Form on Operators. In M. Scheutz (Ed.), *Proceedings of the 4th ACM/IEEE international conference on Human robot interaction* (S. 31–36). New York, NY: ACM.
- Grunwald, A. (2019a). *Der unterlegene Mensch. Die Zukunft der Menschheit im Angesicht von Algorithmen, künstlicher Intelligenz und Robotern* (Originalausgabe, 1. Auflage). München: riva premium.
- Grunwald, A. (2019b, Dezember). *Digitalisierung und künstliche Intelligenz: Gestalten wir die Algorithmen oder gestalten die Algorithmen uns?* IDEepolis 2019, Stuttgart.
- Gunkel, D. J. (2018). The other question: can and should robots have rights? *Ethics and Information Technology*, 20(2), 87–99. <https://doi.org/10.1007/s10676-017-9442-4>

- Gutmann, M. & Knifka, J. (2015). Biomorphic and Technomorphic Metaphors - Some Argument Why Robots Do not Evolve, Why Computing Is not Organic, and Why Adaptive Technologies Are not Intelligent. In M. Decker, M. Gutmann & J. Knifka (Eds.), *Evolutionary robotics, organic computing and adaptive ambience. Epistemological and ethical implications of technomorphic descriptions of technologies* (Hermeneutics and anthropology, Volume/Band 6, S. 53–78). Zürich: LIT.
- Gutmann, M. & Nick, P. (2019). Modellbildung. In P. Nick (Hrsg.), *Modellorganismen* (Lehrbuch, S. 199–242). Berlin, Heidelberg: Springer Berlin Heidelberg.
https://doi.org/10.1007/978-3-662-54868-4_8
- Gutmann, M. & Rathgeber, B. (2010). Notwendige Metaphern? In M. Bölker (Hrsg.), *Information und Menschenbild* (Ethics of Science and Technology Assessment, Bd. 37, S. 173–197). Berlin: Springer. https://doi.org/10.1007/978-3-642-04742-8_10
- Haag, M. (2015). Kollaboratives Arbeiten mit Robotern – Vision und realistische Perspektive. In A. Botthof & E. A. Hartmann (Hrsg.), *Zukunft der Arbeit in Industrie 4.0* (S. 59–64). Berlin, Heidelberg: Springer Berlin Heidelberg.
https://doi.org/10.1007/978-3-662-45915-7_6
- Hancock, P. A., Billings, D. R. & Schaefer, K. E. (2011). Can You Trust Your Robot? *Ergonomics in Design: The Quarterly of Human Factors Applications*, 19(3), 24–29.
<https://doi.org/10.1177/1064804611415045>
- Hancock, P. A., Billings, D. R., Schaefer, K. E., Chen, J. Y. C., de Visser, E. J. & Parasuraman, R. (2011). A meta-analysis of factors affecting trust in human-robot interaction. *Human Factors*, 53(5), 517–527. <https://doi.org/10.1177/0018720811417254>
- Hancock, P. A., Kessler, T. T., Kaplan, A. D., Brill, J. C. & Szalma, J. L. (2020). Evolving Trust in Robots: Specification Through Sequential and Comparative Meta-Analyses. *Human Factors*, 1-34. Online-Vorabveröffentlichung.
<https://doi.org/10.1177/0018720820922080>
- Harari, Y. N. (2015). *Sapiens. A brief history of humankind* (Popular science). London: Vintage Books.
- Hardin, R. (2006). *Trust* (Key concepts in the social sciences). Cambridge: Polity.
- Haslam, N. (2006). Dehumanization: an integrative review. *Personality and Social Psychology Review : an Official Journal of the Society for Personality and Social Psychology, Inc*, 10(3), 252–264. https://doi.org/10.1207/s15327957pspr1003_4
- Hegenberg, J., Schimpf, D. W., Fischer, N. & Schmidt, L. (2019). Pilotstudie zur Roboterunterstützung des Menschen bei manueller Montage. In Gesellschaft für Arbeitswissenschaft (Hrsg.), *GfA-Frühjahrskongress 2019. Arbeit interdisziplinär analysieren - bewerten - gestalten*.
- Held, C., Knauff, M. & Vosgerau, G. (2006). General Introduction: Current Developments in Cognitive Psychology, Neuroscience, and the Philosophy of Mind. In C. Held, M. Knauff & G. Vosgerau (Eds.), *Mental models and the mind. Current*

- developments in cognitive psychology, neuroscience, and philosophy of mind* (Advances in psychology, vol. 138, 1st ed.). Amsterdam: Elsevier.
- Helfferrich, C. (2011). *Die Qualität qualitativer Daten. Manual für die Durchführung qualitativer Interviews* (4. Auflage). Wiesbaden: VS Verlag für Sozialwissenschaften. <https://doi.org/10.1007/978-3-531-92076-4>
- Helfferrich, C. (2019). Leitfaden- und Experteninterviews. In N. Baur & J. Blasius (Hrsg.), *Handbuch Methoden der empirischen Sozialforschung* (2. Aufl. 2019, S. 669–686). Wiesbaden: Springer Fachmedien Wiesbaden. https://doi.org/10.1007/978-3-658-21308-4_44
- Hemforth, B. & Konieczny, L. (2006). Language Processing: Construction of Mental Models or More? In C. Held, M. Knauff & G. Vosgerau (Eds.), *Mental models and the mind. Current developments in cognitive psychology, neuroscience, and philosophy of mind* (Advances in psychology, vol. 138, 1st ed., S. 189–204). Amsterdam: Elsevier.
- Henderson, M. M. (2015). *Industrial Robot Acceptance: Effects of Workforce Demographics and Establishing a Culture of Acceptance within Manufacturing Industry*. Master Thesis. University of South Carolina, Columbia. Zugriff am 27.07.2019. Verfügbar unter: <https://scholarcommons.sc.edu/etd/3235>
- Hergeth, S., Lorenz, L., Krems, J. F. & Toenert, L. (2015). Effects of Take-Over Requests and Cultural Background on Automation Trust in Highly Automated Driving. In *Proceedings of the 8th International Driving Symposium on Human Factors in Driver Assessment, Training, and Vehicle Design: driving assessment 2015* (S. 331–337). Iowa City, Iowa: University of Iowa.
- Herzberg, F., Mausner, B. & Snyderman, B. B. (1959/2011). *The motivation to work*. Transaction Publishers.
- Hesse, H. (1927/2016). *Der Steppenwolf. Erzählung* (Suhrkamp-Taschenbuch, Bd. 175, 56. Auflage). Frankfurt am Main: Suhrkamp.
- Heßler, M. (2015). Die Ersetzung des Menschen? *Technikgeschichte*, 82(2), 108–135. <https://doi.org/10.5771/0040-117X-2015-2-108>
- Heßler, M. (2019). Menschen – Maschinen – MenschMaschinen in Zeit und Raum. Perspektiven einer Historischen Technikanthropologie. In M. Heßler & H. Weber (Hrsg.), *Provokationen der Technikgeschichte. Zum Reflexionszwang historischer Forschung*. Paderborn: Verlag Ferdinand Schöningh. https://doi.org/10.30965/9783657792337_003
- Heyer, C. (2010). Human-robot interaction and future industrial robotics applications. In *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 2010), Taipei, Taiwan, October 18-22, 2010* (S. 4749–4754). Piscataway, N.J.: IEEE.
- Hindemith, L., Vollmer, A.-L., Göpfert, J. P., Wiebel-Herboth, C. B. & Wrede, B. (2020, 5. November). *Why robots should be technical: Correcting mental models through*

- technical architecture concepts*. Preprint (eingereicht bei Interaction Studies Journal). Zugriff am 13.11.2020. Verfügbar unter: <https://ui.adsabs.harvard.edu/abs/2020arXiv201102731H/abstract>
- Hinds, P. J., Roberts, T. L. & Jones, H. (2004). Whose Job Is It Anyway? A Study of Human-Robot Interaction in a Collaborative Task. *Human-Computer Interaction*, (19), 151–181.
- Ho, C.-C. & MacDorman, K. F. (2017). Measuring the Uncanny Valley Effect. *International Journal of Social Robotics*, 9(1), 129–139. <https://doi.org/10.1007/s12369-016-0380-9>
- Hoff, K. A. & Bashir, M. (2015). Trust in automation: integrating empirical evidence on factors that influence trust. *Human Factors*, 57(3), 407–434. <https://doi.org/10.1177/0018720814547570>
- Höffe, O. (Hrsg.). (1997). *Lexikon der Ethik* (Beck'sche Reihe, Bd. 152, Orig.-Ausg., 5., neubearb. und erw. Aufl.). München: Beck.
- Hoffman, G. (2019). Evaluating Fluency in Human–Robot Collaboration. *IEEE Transactions on Human-Machine Systems*, 49(3), 209–218. <https://doi.org/10.1109/THMS.2019.2904558>
- Hoffman, G., Forlizzi, J., Ayal, S., Steinfeld, A., Antanitis, J., Hochman, G. et al. (2015). Robot Presence and Human Honesty: Experimental Evidence. In J. A. Adams, W. D. Smart, B. Mutlu & L. Takayama (Hrsg.), *HRI 2015. Proceedings of the 2015 ACM/IEEE International Conference on Human-Robot Interaction* (S. 181–188). New York, NY: ACM.
- Hofmann, L., Möckelmann, B. & Walach, H. (2003). Entwicklung und empirische Validierung einer Skala zur Erfassung der Einstellung von Psychotherapeuten zum Verhältnis von Psychotherapie und Spiritualität / Religiosität. In W. Belschner, L. Hoffmann & H. Walach (Hrsg.), *Auf dem Weg zu einer Psychologie des Bewusstseins. Transpersonale Studien Band 8* (S. 113-154). Oldenburg: BIS.
- Horstmann, A. C., Bock, N., Linhuber, E., Szczuka, J. M., Straßmann, C. & Krämer, N. C. (2018). Do a robot's social skills and its objection discourage interactants from switching the robot off? *PloS One*, 13(7), Aufsatz e0201581. <https://doi.org/10.1371/journal.pone.0201581>
- Hume, D. (1757/1994). *The natural history of religion* (A library of modern religious thought). Stanford, Calif.: Stanford Univ. Press.
- Ifenthaler, D. & Seel, N. M. (2012). Learning-Dependent Progression of Mental Models. In N. M. Seel (Ed.), *Encyclopedia of the sciences of learning* (Springer reference, S. 2032–2036). New York: Springer. https://doi.org/10.1007/978-1-4419-1428-6_1842
- Im Shin, H. & Kim, J. [Juyoung]. (2018). My computer is more thoughtful than you: Loneliness, anthropomorphism and dehumanization. *Current Psychology*, 27, Aufsatz 1644. <https://doi.org/10.1007/s12144-018-9975-7>

- International Federation of Robotics. (2018). *Demystifying Collaborative Industrial Robots. Positioning Paper*. Zugriff am 06.07.2020. Verfügbar unter: <https://ifr.org/papers/demystifying-collaborative-industrial-robots-updated-version>
- International Federation of Robotics. (2019). *World Robotics Report 2019 - Industrial Robots*. Frankfurt: VDMA Services.
- International Federation of Robotics. (2020a). *Industrial Robots*, International Federation of Robotics. Zugriff am 03.04.2021. Verfügbar unter: <https://ifr.org/industrial-robots/>
- International Federation of Robotics. (2020b). *Service Robots*, International Federation of Robotics. Zugriff am 03.04.2021. Verfügbar unter: <https://ifr.org/service-robots/>
- ISO 8373:2012 (2012-03). *Robots and robotic devices — Vocabulary*.
- Janich, P. (2002). Modelle und Modelliertes. Zwecke und Methoden. In C. F. Gethmann & S. Lingner (Hrsg.), *Integrative Modellierung zum Globalen Wandel* (Wissenschaftsethik und Technikfolgenbeurteilung, Schriftenreihe der Europäischen Akademie zur Erforschung von Folgen wissenschaftlich-technischer Entwicklungen Bad Neuenahr-Ahrweiler GmbH, Bd. 17, S. 15–31). Berlin: Springer.
- Jessup, S. A. & Schneider, T. R. (2021). The role of emotions in human-robot interactions. In C. S. Nam & J. B. Lyons (Eds.), *Trust in Human-Robot Interaction* (S. 515–530). Elsevier Academic Press. <https://doi.org/10.1016/B978-0-12-819472-0.00022-8>
- Jian, J.-Y., Bisantz, A. M. & Drury, C. G. (2000). Foundations for an Empirically Determined Scale of Trust in Automated Systems. *International Journal of Cognitive Ergonomics*, 4(1), 53–71. https://doi.org/10.1207/S15327566IJCE0401_04
- Johnson, D. G. & Verdicchio, M. (2018). Why robots should not be treated like animals. *Ethics and Information Technology*, 20(4), 291–301. <https://doi.org/10.1007/s10676-018-9481-5>
- Johnson-Laird, P. N. (2005). Mental Models and Thinking. In K. J. Holyoak & R. G. Morrison (Eds.), *The Cambridge handbook of thinking and reasoning* (S. 185–208). New York: Cambridge University Press.
- Johnson-Laird, P. N. (2010). Mental models and human reasoning. *Proceedings of the National Academy of Sciences of the United States of America*, 107(43), 18243–18250. <https://doi.org/10.1073/pnas.1012933107>
- Johnson-Laird, P. N. & Khemlani, S. S. (2017). Mental Models and Causation. *Oxford Handbook of Causal Reasoning*.
- Jones, N. A., Ross, H., Lynam, T. & Perez, P. (2014). Eliciting Mental Models: a Comparison of Interview Procedures in the Context of Natural Resource Management. *Ecology and Society*, 19(1), Aufsatz 13. <https://doi.org/10.5751/ES-06248-190113>
- Jones, N. A., Ross, H., Lynam, T., Perez, P. & Leitch, A. (2011). Mental Models: An Interdisciplinary Synthesis of Theory and Methods. *Ecology and Society*, 16(1), Aufsatz 46.

- Jung, M. F. & Hinds, P. J. (2018). Robots in the Wild: A Time for More Robust Theories of Human-Robot Interaction. *ACM Trans. Hum.-Robot Interact.*, 7(1), Aufsatz 2. <https://doi.org/10.1145/3208975>
- Juvina, I., Collins, M. G., Larue, O., Kennedy, W. G., de Visser, E. J. & Melo, C. D. (2019). Toward a Unified Theory of Learned Trust in Interpersonal and Human-Machine Interactions. *ACM Transactions on Interactive Intelligent Systems*, 9(4), 1–33. <https://doi.org/10.1145/3230735>
- Kahn, P. H., Kanda, T., Ishiguro, H., Freier, N. G., Severson, R. L., Gill, B. T. et al. (2012). "Robovie, you'll have to go into the closet now". Children's social and moral relationships with a humanoid robot. *Developmental Psychology*, 48(2), 303–314. <https://doi.org/10.1037/a0027033>
- Kahneman, D. & Tversky, A. (1984). Choices, Values, and Frames. *American Psychologist*, 39(4), 341–350. <https://doi.org/10.1017/CBO9780511803475.002>
- Kamide, H. & Arai, T. (2017). Perceived Comfortableness of Anthropomorphized Robots in U.S. and Japan. *International Journal of Social Robotics*, 9(4), 537–543. <https://doi.org/10.1007/s12369-017-0409-8>
- Kant, I. (1924/1991). *Eine Vorlesung über Ethik* (Fischer-Taschenbücher, Bd. 10249, Neuauflg.). Frankfurt am Main: Fischer-Taschenbuch-Verl.
- Kaplan, A. D., Kessler, T. T., Sanders, T. L., Cruik, J., Brill, J. C. & Hancock, P. A. (2021). A time to trust: Trust as a function of time in human-robot interaction. In C. S. Nam & J. B. Lyons (Eds.), *Trust in Human-Robot Interaction* (S. 143–157). Elsevier Academic Press. <https://doi.org/10.1016/B978-0-12-819472-0.00006-X>
- Kaplan, A. D., Sanders, T. L. & Hancock, P. A. (2021). Likert or Not? How Using Likert Rather Than Bipolar Ratings Reveal Individual Difference Scores Using the God-speed Scales. *International Journal of Social Robotics*. Online-Vorabveröffentlichung. <https://doi.org/10.1007/s12369-020-00740-y>
- Kawakami, N. (2012). The implicit influence of a negative mood on the subliminal mere exposure effect. *Perceptual and Motor Skills*, 115(3), 715–724. <https://doi.org/10.2466/22.24.27.PMS.115.6.715-724>
- Kehl, C. & Coenen, C. (2016). *Technologien und Visionen der Mensch-Maschine-Entgrenzung*. Büro für Technikfolgen-Abschätzung beim Deutschen Bundestag (TAB). Zugriff am 06.03.2020. Verfügbar unter: <https://www.tab-beim-bundestag.de/de/pdf/publikationen/berichte/TAB-Arbeitsbericht-ab167.pdf>
- Kelle, U. (2019). Mixed Methods. In N. Baur & J. Blasius (Hrsg.), *Handbuch Methoden der empirischen Sozialforschung* (2. Aufl. 2019, S. 159–172). Wiesbaden: Springer Fachmedien Wiesbaden. https://doi.org/10.1007/978-3-658-21308-4_10
- Kessler, T. T., Stowers, K., Brill, J. C. & Hancock, P. A. (2017). Comparisons of Human-Human Trust with Other Forms of Human-Technology Trust. In *Proceedings of the Human Factors and Ergonomics Society 2017 Annual Meeting* (S. 1303–1307).

- Khavas, Z. R., Ahmadzadeh, S. R. & Robinette, P. (2020). Modeling Trust in Human-Robot Interaction: A Survey. In A. R. Wagner, D. Feil-Seifer & K. S. Haring (Eds.), *Social Robotics. 12th International Conference, ICSR 2020, Golden, CO, USA, November 14–18, 2020, Proceedings* (Lecture Notes in Artificial Intelligence, 1st ed., vol. 12483, pp. 529–541). https://doi.org/10.1007/978-3-030-62056-1_44
- Kiesler, S. (2005). Fostering common ground in human-robot interaction. In *2005 IEEE International Workshop on Robot and Human Interactive Communication (RO-MAN), Nashville, TN, August 13-15, 2005* (S. 729–734). Piscataway, N.J: IEEE.
- Kiesler, S. & Goetz, J. (2002). Mental Models and Cooperation with Robotic Assistants. *CHI '02 Extended Abstracts on Human Factors in Computing Systems*, 576–577. <https://doi.org/10.1145/506443.506491>
- Kim, B., Bruce, M., Brown, L., de Visser, E. J. & Phillips, E. K. (2020). A Comprehensive Approach to Validating the Uncanny Valley using the Anthropomorphic RoBOT (ABOT) Database. In *2020 Systems and Information Engineering Design Symposium (SIEDS), University of Virginia (Virtual Conference), Charlottesville, Virginia, USA, 24 April 2020* (S. 1–6). Piscataway, New Jersey: IEEE.
- Kim, W., Kim, N., Lyons, J. B. & Nam, C. S. (2020). Factors affecting trust in high-vulnerability human-robot interaction contexts: A structural equation modelling approach. *Applied Ergonomics*, 85, Aufsatz 103056. <https://doi.org/10.1016/j.apergo.2020.103056>
- Kim, Y. & Sundar, S. S. (2012). Anthropomorphism of computers: Is it mindful or mindless? *Computers in Human Behavior*, 28(1), 241–250. <https://doi.org/10.1016/j.chb.2011.09.006>
- Klapper, A., Ramsey, R., Wigboldus, D. & Cross, E. S. (2014). The control of automatic imitation based on bottom-up and top-down cues to animacy: insights from brain and behavior. *Journal of Cognitive Neuroscience*, 26(11), 2503–2513. https://doi.org/10.1162/jocn_a_00651
- Knapp, T. R. & Schafer, W. D. (2009). From Gain Score t to ANCOVA F (and vice versa). *Practical Assessment, Research & Evaluation*, 14, Aufsatz 6. <https://doi.org/10.7275/YKE1-K937>
- Knauff, M. (2006). A Neuro-Cognitive Theory of Relational Reasoning with Mental Models and Visual Images. In C. Held, M. Knauff & G. Vosgerau (Eds.), *Mental models and the mind. Current developments in cognitive psychology, neuroscience, and philosophy of mind* (Advances in psychology, vol. 138, 1st ed., S. 127–152). Amsterdam: Elsevier.
- Knauff, M. & Johnson-Laird, P. N. (2002). Visual imagery can impede reasoning. *Memory & Cognition*, 30(3), 363–371. <https://doi.org/10.3758/bf03194937>
- Koch, T., Peter, C. & Müller, P. (2019). *Das Experiment in der Kommunikations- und Medienwissenschaft*. Wiesbaden: Springer Fachmedien Wiesbaden. <https://doi.org/10.1007/978-3-658-19754-4>

- Komatsu, T., Kurosawa, R. & Yamada, S. (2012). How Does the Difference Between Users' Expectations and Perceptions About a Robotic Agent Affect Their Behavior? *International Journal of Social Robotics*, 4(2), 109–116.
<https://doi.org/10.1007/s12369-011-0122-y>
- Kompatsiari, K., Pérez-Osorio, J., Tommaso, D. de, Metta, G. & Wykowska, A. (2018). Neuroscientifically-Grounded Research for Improved Human-Robot Interaction. In *2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)* (S. 3403–3408).
- Kopp, T. (2020). Die Wirkkraft der Sprache auf die Wahrnehmung neuartiger Technologien wie kollaborationsfähiger Roboter (Cobots) oder künstlicher Intelligenz (KI). In F. Nees, I. Stengel, V. G. Meister, T. Barton, F. Herrmann, C. Müller et al. (Hrsg.), *Angewandte Forschung in der Wirtschaftsinformatik 2020. Tagungsband zur 33. AKWI-Jahrestagung am 14.09.2020* (S. 41–52). Heide: mana Buch.
- Kopp, T., Baumgartner, M. & Kinkel, S. (2021). Success factors for introducing industrial human-robot interaction in practice: an empirically driven framework. *The International Journal of Advanced Manufacturing Technology*, 112(3-4), 685–704.
<https://doi.org/10.1007/s00170-020-06398-0>
- Kopp, T., Baumgartner, M. & Kinkel, S. (2022). How Linguistic Framing Affects Factory Workers' Initial Trust in Collaborative Robots: The Interplay Between Anthropomorphism and Technological Replacement. *International Journal of Human-Computer Studies*, 158, Aufsatz 102730. <https://doi.org/10.1016/j.ijhcs.2021.102730>
- Kopp, T., Hendig, I. & Kinkel, S. (2021). Kollaborationsfähige Roboter: Hype oder Zukunftstechnologie? Ergebnisse einer empirischen Marktanalyse. *Industrie 4.0 Management*, (3), 53–57.
- Kopp, T., Schäfer, A. & Kinkel, S. (2020). Kollaborierende oder kollaborationsfähige Roboter? Welche Rolle spielt die Mensch-Roboter-Kollaboration in der Praxis? *Industrie 4.0 Management*, 2020(2), 19–23. https://doi.org/10.30844/I40M_20-2_S19-23
- Kornmeier, M. (2018). *Wissenschaftlich schreiben leicht gemacht. Für Bachelor, Master und Dissertation* (UTB Schlüsselkompetenzen, Bd. 3154, 8., überarbeitete Auflage). Bern: Haupt Verlag.
- Kory, J. & Kleinberger, R. (2014). Social agent or machine? An exploration of how the framing of a robot affects prosodic mimicry and expressivity. *2nd Workshop on Applications for Emotional Robots held in conjunction with the 9th ACM/IEEE International Conference on Human-Robot Interaction*.
- Kory Westlund, J. M. & Breazeal, C. L. (2019). Exploring the Effects of a Social Robot's Speech Entrainment and Backstory on Young Children's Emotion, Rapport, Relationship, and Learning. *Frontiers in Robotics and AI*, 6, Aufsatz 54.
<https://doi.org/10.3389/frobt.2019.00054>

- Kory Westlund, J. M., Jeong, S., Park, H. W., Ronfard, S., Adhikari, A., Harris, P. L. et al. (2017). Flat vs. Expressive Storytelling: Young Children's Learning and Retention of a Social Robot's Narrative. *Frontiers in Human Neuroscience*, *11*, Aufsatz 295. <https://doi.org/10.3389/fnhum.2017.00295>
- Kory Westlund, J. M., Martinez, M., Archie, M., Das, M. & Breazeal, C. L. (2016). Effects of Framing a Robot as a Social Agent or as a Machine on Children's Social Behavior. In *RO-MAN 2016. The 25th IEEE International Symposium on Robot and Human Interactive Communication* (S. 688–693). Piscataway, NJ: IEEE.
- Krägeloh, C. U., Bharatharaj, J., Sasthan Kutty, S. K., Nirmala, P. R. & Huang, L. (2019). Questionnaires to Measure Acceptability of Social Robots: A Critical Review. *Robotics*, *8*(4), Aufsatz 88. <https://doi.org/10.3390/robotics8040088>
- Kramer, R. M. (1999). Trust and distrust in organizations: emerging perspectives, enduring questions. *Annual Review of Psychology*, *50*, 569–598. <https://doi.org/10.1146/annurev.psych.50.1.569>
- Krippendorff, K. (1970). Bivariate Agreement Coefficients for Reliability of Data. *Sociological Methodology*, *2*, 139–150. <https://doi.org/10.2307/270787>
- Krüger, J., Lien, T. K. & Verl, A. (2009). Cooperation of human and machines in assembly lines. *CIRP Annals*, *58*(2), 628–646. <https://doi.org/10.1016/j.cirp.2009.09.009>
- Kuchenbrandt, D. & Eyssel, F. (2012). The mental simulation of a human-robot interaction: Positive effects on attitudes and anxiety toward robots. In *2012 IEEE Ro-Man. The 21st IEEE International Symposium on Robot and Human Interactive Communication* (S. 463–468). IEEE.
- Kuckartz, U. (2018). *Qualitative Inhaltsanalyse. Methoden, Praxis, Computerunterstützung* (Grundlagentexte Methoden, 4. Auflage). Weinheim, Basel: Beltz Juventa.
- Kuz, S., Mayer, M. P., Müller, S. & Schlick, C. M. (2017). Using Anthropomorphism to Improve the Human-Machine Interaction in Industrial Environments (Part I). In V. G. Duffy (Ed.), *Digital Human Modeling. Applications in Health, Safety, Ergonomics, and Risk Management: Ergonomics and Design. Proceedings of the 8th International Conference, DHM 2017, Held as Part of HCI International, Vancouver, BC, Canada, July 9-14, 2017, Part I* (vol. 8026, S. 76–85). Cham: Springer International Publishing. https://doi.org/10.1007/978-3-642-39182-8_9
- Lammer, L., Huber, A., Weiss, A. & Vincze, M. (2014). Mutual Care: How older adults react when they should help their care robot. In *40th Annual Convention of the Society for the Study of Artificial Intelligence and Simulation of Behaviour 2014 (AISB 2014)* (S. 363–370). Red Hook, NY: Curran.
- Lantz, B. (2013). The large sample size fallacy. *Scandinavian Journal of Caring Sciences*, *27*(2), 487–492. <https://doi.org/10.1111/j.1471-6712.2012.01052.x>
- Laue, C. (2017). Familiar and Strange: Gender, Sex, and Love in the Uncanny Valley. *Multimodal Technologies and Interaction*, *1*(1), 1–11. <https://doi.org/10.3390/mti1010002>

- Law, T., Chita-Tegmark, M. & Scheutz, M. (2021). The Interplay Between Emotional Intelligence, Trust, and Gender in Human–Robot Interaction. *International Journal of Social Robotics*, 13(2), 297–309. <https://doi.org/10.1007/s12369-020-00624-1>
- Law, T., Malle, B. F. & Scheutz, M. (2021). A Touching Connection: How Observing Robotic Touch Can Affect Human Trust in a Robot. *International Journal of Social Robotics*. Online-Vorabveröffentlichung. <https://doi.org/10.1007/s12369-020-00729-7>
- Law, T. & Scheutz, M. (2021). Trust: Recent concepts and evaluations in human-robot interaction. In C. S. Nam & J. B. Lyons (Eds.), *Trust in Human-Robot Interaction* (S. 27–57). Elsevier Academic Press. <https://doi.org/10.1016/B978-0-12-819472-0.00002-2>
- Lecheler, S. (2020). Dan, Viorela: Integrative framing analysis: framing health through words and visuals. *Publizistik*, 65(4), 663–664. <https://doi.org/10.1007/s11616-020-00604-5>
- Lecheler, S. & de Vreese, C. H. (2011). Getting Real: The Duration of Framing Effects. *Journal of Communication*, 61(5), 959–983. <https://doi.org/10.1111/j.1460-2466.2011.01580.x>
- Lecheler, S. & de Vreese, C. H. (2016). How Long Do News Framing Effects Last? A Systematic review of Longitudinal Studies. *Annals of the International Communication Association*, 40(1), 3–30. <https://doi.org/10.1080/23808985.2015.11735254>
- Lee, J. D. & See, K. A. (2004). Trust in automation: designing for appropriate reliance. *Human Factors: the Journal of the Human Factors and Ergonomics Society*, 46(1), 50–80. https://doi.org/10.1518/hfes.46.1.50_30392
- Lee, M. K., Kiesler, S. & Forlizzi, J. (2010). Receptionist or information kiosk. In K. Inkpen (ed.), *Proceedings of the 2010 ACM conference on Computer supported cooperative work* (S. 31–40). New York, NY: ACM.
- Lee, M. K., Kiesler, S., Forlizzi, J. & Rybski, P. (2012). Ripple effects of an embedded social agent. A field study of a social robot in the workplace. In J. A. Konstan, E. H. Chi & K. Höök (eds.), *Proceedings of the 2012 ACM annual conference on Human Factors in Computing Systems - CHI '12* (S. 695–704). New York, NY, USA: ACM Press.
- Lee, S., Kiesler, S., Lau, I. Y. & Chiu, C.-Y. (2005). Human Mental Models of Humanoid Robots. In *2005 IEEE International Conference on Robotics and Automation (ICRA). Barcelona, Spain: April 18-22 2005* (S. 2767–2772). Piscataway, N.J.: IEEE.
- Lefkeli, D., Ozbay, Y., Gürhan-Canli, Z. & Eskenazi, T. (2020). Competing with or Against Cozmo, the Robot: Influence of Interaction Context and Outcome on Mind Perception. *International Journal of Social Robotics*. Online-Vorabveröffentlichung. <https://doi.org/10.1007/s12369-020-00668-3>
- Legler, F., Langer, D., Dittrich, F. & Bullinger, A. C. (2020). I don't care what the robot does! Trust in automation when working with a heavy-load robot. In D. de Waard, A. Toffetti, L. Pietrantoni, T. Franke, J.-F. Petiot, C. Dumas et al. (Hrsg.), *Proceedings*

- of the Human Factors and Ergonomics Society Europe Chapter 2019 Annual Conference* (S. 239–253).
- Leifeld, P. & Haunss, S. (2012). Political discourse networks and the conflict over software patents in Europe. *European Journal of Political Research*, 51(3), 382–409. <https://doi.org/10.1111/j.1475-6765.2011.02003.x>
- Lemaignan, S., Fink, J. & Dillenbourg, P. (2014). The dynamics of anthropomorphism in robotics. In G. Sagerer, M. Imai, T. Belpaeme & A. L. Thomaz (eds.), *HRI 2014. Proceedings of the 2014 ACM/IEEE International Conference on Human-Robot Interaction* (S. 226–227). New York, NY: ACM.
- Lemaignan, S., Fink, J., Dillenbourg, P. & Braboszcz, C. (2014). The Cognitive Correlates of Anthropomorphism. *Human-Robot Interaction Conference, Workshop "HRI: a bridge between Robotics and Neuroscience"*.
- Lenz, C. (2011). *Context-aware human-robot collaboration as a basis for future cognitive factories. Dissertation*. Technische Universität München, München. Zugriff am 15.11.2017. Verfügbar unter: <http://mediatum.ub.tum.de/doc/1287170/document.pdf>
- Letheren, K., Kuhn, K.-A. L., Lings, I. & Pope, N. K. L. (2016). Individual difference factors related to anthropomorphic tendency. *European Journal of Marketing*, 50(5/6), 973–1002. <https://doi.org/10.1108/EJM-05-2014-0291>
- Levin, D., Harriott, C., Paul, N. A., Zheng, T. & Adams, J. A. (2013). Cognitive Dissonance as a Measure of Reactions to Human-Robot Interaction. *Journal of Human-Robot Interaction*, 2(3), 1–17. <https://doi.org/10.5898/JHRI.2.3.Levin>
- Lévinas, E. & Krewani, W. N. (1993). *Totalität und Unendlichkeit. Versuch über die Exteriorität* (Alber-Reihe Philosophie, 2., unveränd. Aufl.). Freiburg: Alber.
- Levy, D. (2017). Why Not Marry a Robot? In A. D. Cheok, K. Devlin & D. Levy (Eds.), *Love and sex with robots. Second international conference, LSR 2016, London, UK, december 19-20, 2016 : revised selected papers* (Lecture notes in computer science Lecture notes in artificial intelligence, vol. 10237, S. 3–13). Cham: Springer. https://doi.org/10.1007/978-3-319-57738-8_1
- Lewicki, R. J., McAllister, D. J. & Bies, R. J. (1998). Trust And Distrust: New Relationships and Realities. *Academy of Management Review*, 23(3), 438–458. <https://doi.org/10.5465/amr.1998.926620>
- Lewicki, R. J., Tomlinson, E. C. & Gillespie, N. (2006). Models of Interpersonal Trust Development: Theoretical Approaches, Empirical Evidence, and Future Directions. *Journal of Management*, 32(6), 991–1022. <https://doi.org/10.1177/0149206306294405>
- Lewis, M., Sycara, K. & Walker, P. (2018). The Role of Trust in Human-Robot Interaction. In H. A. Abbass, J. Scholz & D. J. Reid (Eds.), *Foundations of Trusted Autonomy* (vol. 117, S. 135–160). Cham: Springer International Publishing.

- Li, D., Rau, P. L. P. & Li, Y. (2010). A Cross-cultural Study. Effect of Robot Appearance and Task. *International Journal of Social Robotics*, 2(2), 175–186.
<https://doi.org/10.1007/s12369-010-0056-9>
- Litzenberger, G. (2019). *IFR publishes collaborative industrial robot definition and estimates supply*, International Federation of Robotics. IFR Secretariat Blog. Zugriff am 24.06.2020. Verfügbar unter: <https://ifr.org/post/international-federation-of-robotics-publishes-collaborative-industrial-rob>
- Ljungblad, S., Kotrbova, J., Jacobsson, M., Cramer, H. & Niechwiadowicz, K. (2012). Hospital Robot at Work: Something Alien or an Intelligent Colleague? In *Proceedings of the ACM 2012 conference on Computer Supported Cooperative Work* (S. 177–186). Zugriff am 12.04.2018.
- Loh, J. (2019). *Roboterethik. Eine Einführung* (suhrkamp taschenbuch wissenschaft, Erste Auflage, Originalausgabe). Berlin: Suhrkamp.
- Lohse, M. (2009). *Investigating the influence of situations and expectations on user behavior - empirical analyses in human-robot interaction*. Dissertation. Universität Bielefeld, Bielefeld. Zugriff am 09.05.2018. Verfügbar unter: https://pub.uni-bielefeld.de/download/2303768/2303771/Thesis_Lohse.pdf
- Lohse, M. (2012). Treating robots as social beings - A matter of personal preconceptions or interpersonal alignment? In *2012 IEEE Ro-Man. The 21st IEEE International Symposium on Robot and Human Interactive Communication* (S. 839–844). IEEE.
- Lombardi, C. M. & Hurlbert, S. H. (2009). Misprescription and misuse of one-tailed tests. *Austral Ecology*, 34, 447–468. <https://doi.org/10.1111/j.1442-9993.2009.01946.x>
- Lord, F. M. (1967). A Paradox in the Interpretation of Group Comparisons. *Psychological Bulletin*, 68(5), 304–305.
- Luhmann, N. (1968/2014). *Vertrauen. Ein Mechanismus der Reduktion sozialer Komplexität* (5. Aufl.). UTB.
- Lyons, J. B. & Guznov, S. Y. (2019). Individual differences in human-machine trust: A multi-study look at the perfect automation schema. *Theoretical Issues in Ergonomics Science*, 20(4), 440–458. <https://doi.org/10.1080/1463922X.2018.1491071>
- Lyons, J. B. & Nam, C. S. (2021). Introduction: The evolution of trust in human-robot interaction. In C. S. Nam & J. B. Lyons (Eds.), *Trust in Human-Robot Interaction* (S. xxi–xxv). Elsevier Academic Press.
- Lyons, J. B. & Wynne, K. T. (2021). Human-machine teaming: Evaluating dimensions using narratives. *Human-Intelligent Systems Integration*. Online-Vorabveröffentlichung. <https://doi.org/10.1007/s42454-020-00019-7>
- Lyons, J. B., Wynne, K. T., Mahoney, S. & Roebke, M. A. (2019). Trust and Human-Machine Teaming: A Qualitative Study. In W. F. Lawless, I. S. Moskowitz, R. Mittu, S. Russell & D. Sofge (Eds.), *Artificial Intelligence for the Internet of Everything* (S. 101–116). Academic Press. <https://doi.org/10.1016/B978-0-12-817636-8.00006-5>

- Mabry, L. (2016). *The Rhetoric of Social Robots: How Consumerism is Shaping Perceptions of Robotic Ontology*. Master Thesis. University of Washington, Seattle, Washington, USA.
- MacDorman, K. F., Vasudevan, S. K. & Ho, C.-C. (2009). Does Japan really have robot mania? Comparing attitudes by implicit and explicit measures. *AI & SOCIETY*, 23(4), 485–510. <https://doi.org/10.1007/s00146-008-0181-2>
- Madhavan, P. & Wiegmann, D. A. (2007). Similarities and differences between human–human and human–automation trust: an integrative review. *Theoretical Issues in Ergonomics Science*, 8(4), 277–301. <https://doi.org/10.1080/14639220500337708>
- Majdandžić, J., Bauer, H., Windischberger, C., Moser, E., Engl, E. & Lamm, C. (2012). The human factor: behavioral and neural correlates of humanized perception in moral decision making. *PLoS One*, 7(10), Aufsatz e47698. <https://doi.org/10.1371/journal.pone.0047698>
- Malik, A. A. & Bilberg, A. (2017). Framework to Implement Collaborative Robots In Manual Assembly: A Lean Automation Approach. In B. Katalinic (Ed.), *Proceedings of the 28th International DAAAM Symposium 2017* (DAAAM Proceedings, vol. 1, S. 1151–1160). DAAAM International Vienna. <https://doi.org/10.2507/28th.daaam.proceedings.160>
- Malik, A. A. & Bilberg, A. (2019). Developing a reference model for human–robot interaction. *International Journal on Interactive Design and Manufacturing (IJIDeM)*, 13(5), 1541–1547. <https://doi.org/10.1007/s12008-019-00591-6>
- Malle, B. F., Fischer, K., Young, J. E., Moon, A. & Collins, E. C. (2020). Trust and the Discrepancy between Expectations and Actual Capabilities of Social Robots. In D. Zhang & B. Wei (Eds.), *Human-Robot Interaction. Control, Analysis, and Design* (S. 1–23). Newcastle-upon-Tyne: Cambridge Scholars Publisher.
- Malle, B. F. & Ullman, D. (2021). A multi-dimensional conception and measure of human-robot trust. In C. S. Nam & J. B. Lyons (Eds.), *Trust in Human-Robot Interaction* (S. 3–25). Elsevier Academic Press.
- Marikyan, D., Papagiannidis, S. & Alamanos, E. (2020). Cognitive Dissonance in Technology Adoption: A Study of Smart Home Users. *Information Systems Frontiers: a Journal of Research and Innovation*. Online-Vorabveröffentlichung. <https://doi.org/10.1007/s10796-020-10042-3>
- Marquardt, M. (2017). Anthropomorphisierung in der Mensch-Roboter Interaktionsforschung: theoretische Zugänge und soziologisches Anschlusspotential. *Working Papers - kultur- und techniksoziologische Studien*, 10(1), 1–44. Zugriff am 03.04.2020. Verfügbar unter: <https://nbn-resolving.org/urn:nbn:de:0168-ssoar-57037-3>
- Marquie, J. C., Jourdan-Boddaert, L. & Huet, N. (2002). Do older adults underestimate their actual computer knowledge? *Behaviour & Information Technology*, 21(4), 273–280. <https://doi.org/10.1080/0144929021000020998>

- Marsh, S. & Dibben, M. R. (2003). The role of trust in information science and technology. *Annual Review of Information Science and Technology*, 37(1), 465–498. <https://doi.org/10.1002/aris.1440370111>
- Matthews, G., Panganiban, A. R., Lin, J., Long, M. D. & Schwing, M. (2021). Super-machines or sub-humans: Mental models and trust in intelligent autonomous systems. In C. S. Nam & J. B. Lyons (Eds.), *Trust in Human-Robot Interaction* (S. 59–82). Elsevier Academic Press. <https://doi.org/10.1016/B978-0-12-819472-0.00003-4>
- Matthias, B. & Ding, H. (2013). Die Zukunft der Mensch-Roboter Kollaboration in der industriellen Montage. *Conference: Internationales Forum Mechatronics (ifm) 2013*.
- Maurtua, I., Ibarguren, A., Kildal, J., Susperregi, L. & Sierra, B. (2017). Human–robot collaboration in industrial applications. *International Journal of Advanced Robotic Systems*, 14(4), 1-10. <https://doi.org/10.1177/1729881417716010>
- May, D. C., Holler, K. J., Bethel, C. L., Strawderman, L., Carruth, D. W. & Usher, J. M. (2017). Survey of Factors for the Prediction of Human Comfort with a Non-anthropomorphic Robot in Public Spaces. *International Journal of Social Robotics*, 9(2), 165–180. <https://doi.org/10.1007/s12369-016-0390-7>
- Mayer, R. C., Davis, J. H. & Schoorman, F. D. (1995). An Integrative Model of Organizational Trust. *The Academy of Management Review*, 20(3), 709–734.
- Mayring, P. (2015). *Qualitative Inhaltsanalyse. Grundlagen und Techniken* (12., überarbeitete Auflage). Weinheim, Basel: Beltz Verlag.
- Mayring, P. & Fenzl, T. (2019). Qualitative Inhaltsanalyse. In N. Baur & J. Blasius (Hrsg.), *Handbuch Methoden der empirischen Sozialforschung* (2. Aufl. 2019, S. 633–648). Wiesbaden: Springer Fachmedien Wiesbaden. https://doi.org/10.1007/978-3-658-21308-4_42
- McCarthy, N. (2019). *Infographic: The Countries Facing The Greatest Skill Shortages*, Statistisches Bundesamt. Zugriff am 24.06.2020. Verfügbar unter: <https://www.statista.com/chart/4690/the-countries-facing-the-greatest-skill-shortages/>
- McKnight, D. H. & Chervany, N. L. (2001). Trust and Distrust Definitions: One Bite at a Time. In R. Falcone, M. Singh & Y.-H. Tan (Eds.), *Trust in Cyber-societies. Integrating the Human and Artificial Perspectives* (Lecture Notes in Computer Science, vol. 2246, S. 27–54). Berlin: Springer. https://doi.org/10.1007/3-540-45547-7_3
- Meißner, A. & Trübswetter, A. (2018). Mensch-Roboter-Kollaboration in der Produktion: Kritische Würdigung etablierter Technikakzeptanzmodelle und neue Erkenntnisse in der Akzeptanzforschung. In R. Weidner & A. Karafillidis (eds.), *Technische Unterstützungssysteme, die die Menschen wirklich wollen. Dritte transdisziplinäre Konferenz: Hamburg 2018* (S. 223–233). Hamburg, Deutschland: Helmut-Schmidt-Universität.
- Meißner, A., Trübswetter, A., Conti-Kufner, A. S. & Schmidtler, J. (2020). Friend or Foe? Understanding Assembly Workers' Acceptance of Human-robot Collaboration.

- ACM Transactions on Human-Robot Interaction*, 10(1), 1–30.
<https://doi.org/10.1145/3399433>
- Melson, G. F., Kahn, P. H., Beck, A. M., Friedman, B., Roberts, T. & Garrett, E. (2005). Robots as Dogs? – Children’s Interactions with the Robotic Dog AIBO and a Live Australian Shepherd. *CHI’05 extended abstracts on Human factors in computing systems*, 1649–1652.
- Merenkov, A. V., Campa, R. & Dronishinets, N. P. (2021). Public Opinion on Artificial Intelligence Development. *KnE Social Sciences*, 565–574.
<https://doi.org/10.18502/kss.v5i2.8401>
- Merrill, M. D. (2000). Knowledge objects and mental models. In *Proceedings International Workshop on Advanced Learning Technologies. IWALT 2000. Advanced Learning Technology: Design and Development Issues* (S. 244–246).
- Merritt, S. M., Unnerstall, J. L., Lee, D. & Huber, K. (2015). Measuring Individual Differences in the Perfect Automation Schema. *Human Factors*, 57(5), 740–753.
<https://doi.org/10.1177/0018720815581247>
- Messariss, P. (2000). Book Review: The Media Equation: How People Treat Computers, Television, and New Media Like Real People and Places. *Philosophy of the Social Sciences*, 30(1), 120–124. <https://doi.org/10.1177/004839310003000107>
- Michael, J. & Salice, A. (2017). The Sense of Commitment in Human-Robot Interaction. *International Journal of Social Robotics*, 9(5), 755–763.
<https://doi.org/10.1007/s12369-016-0376-5>
- Miles, E. & Crisp, R. J. (2014). A meta-analytic test of the imagined contact hypothesis. *Group Processes & Intergroup Relations*, 17(1), 3–26.
<https://doi.org/10.1177/1368430213510573>
- Miller, C. A. (2021). Trust, transparency, explanation, and planning: Why we need a lifecycle perspective on human-automation interaction. In C. S. Nam & J. B. Lyons (Eds.), *Trust in Human-Robot Interaction* (S. 233–257). Elsevier Academic Press.
<https://doi.org/10.1016/B978-0-12-819472-0.00011-3>
- Misselhorn, C. (2013). Robots as Moral Agents? In F. Rövekamp & F. Bosse (Eds.), *Ethics in science and society: German and Japanese views* (S. 42–56). München: Iudicium.
- Möllering, G. (2006). *Das Aufheben von Ungewissheit als Kern des Vertrauens: Just do it?* MPIfG Working Paper 5. Köln: Max-Planck-Institut für Gesellschaftsforschung. Zugriff am 17.09.2020.
- Moniz, A. (2014). Organisational Challenges of Human– Robot Interaction Systems in Industry: Human Resources Implications. In C. Machado & J. P. Davim (Eds.), *Human Resource Management and Technological Challenges* (S. 123–131). Springer.
- Moniz, A., Fischer, M. & Krings, B.-J. (2017). Herausforderungen der Mensch-Roboter-Kollaboration. *lernen & lehren*, (1), 8–14.

- Moniz, A. & Krings, B.-J. (2014). Technology assessment approach to human-robot interactions in work environments. In *2014 7th International Conference on Human System Interactions (HSI)* (S. 282–289).
- Moniz, A. & Krings, B.-J. (2016). Robots Working with Humans or Humans Working with Robots? Searching for Social Dimensions in New Human-Robot Interaction in Industry. *Societies*, 6(3), 1–21. <https://doi.org/10.3390/soc6030023>
- Montoya, R. M., Horton, R. S., Vevea, J. L., Citkowicz, M. & Lauber, E. A. (2017). A re-examination of the mere exposure effect: The influence of repeated exposure on recognition, familiarity, and liking. *Psychological Bulletin*, 143(5), 459–498. <https://doi.org/10.1037/bul0000085>
- Mori, M. (1970). The Uncanny Valley. *Energy*, 7, 33–35.
- Mou, Y. & Xu, K. (2017). The media inequality: Comparing the initial human-human and human-AI social interactions. *Computers in Human Behavior*, 72, 432–440. <https://doi.org/10.1016/j.chb.2017.02.067>
- Muir, B. M. & Moray, N. (1996). Trust in automation. Part II. Experimental studies of trust and human intervention in a process control simulation. *Ergonomics*, 39(3), 429–460. <https://doi.org/10.1080/00140139608964474>
- Müller, S. L. & Richert, A. (2018). The Big-Five Personality Dimensions and Attitudes to-wards Robots. In *Proceedings of the 11th Pervasive Technologies Related to Assis-tive Environments Conference - PETRA '18* (S. 405–408). New York, New York, USA: ACM Press.
- Müller, S. L., Schröder, S., Jeschke, S. & Richert, A. (2017). Design of a Robotic Work-mate. In V. G. Duffy (Ed.), *Digital Human Modeling. Applications in Health, Safety, Ergonomics, and Risk Management: Ergonomics and Design. Proceedings of the 8th International Conference, DHM 2017, Held as Part of HCI International, Vancouver, BC, Canada, July 9-14, 2017, Part I* (S. 447–456). Cham: Springer International Pub-lishing. https://doi.org/10.1007/978-3-319-58463-8_37
- Müller, S. L., Shehadeh, M. A., Schröder, S., Richert, A. & Jeschke, S. (2018). An over-view of work analysis instruments for hybrid production workplaces. *AI & SOCIETY*, 33(3), 425–432. <https://doi.org/10.1007/s00146-017-0757-9>
- Müller-Abdelrazeq, S. L. (2020). *Triadisches Phasenmodell des Zusammenspiels von Einstellungssubjekt, -objekt und -kontext bei der industriellen Mensch-Roboter-Inter-aktion*. Aachen: Apprimus Verlag.
- Müller-Abdelrazeq, S. L., Schönefeld, K., Haberstroh, M. & Hees, F. (2019). Interacting with Collaborative Robots—A Study on Attitudes and Acceptance in Industrial Con-texts. In O. Korn (Ed.), *Social Robots. Technological, societal and ethical aspects of human-robot* (Human-Computer Interaction Series, vol. 4, S. 101–117). Berlin: Springer Nature. https://doi.org/10.1007/978-3-030-17107-0_6

- Murashov, V., Hearl, F. & Howard, J. (2016). Working safely with robot workers: Recommendations for the new workplace. *Journal of Occupational and Environmental Hygiene*, 13(3), D61-D71. <https://doi.org/10.1080/15459624.2015.1116700>
- Mutlu, B. & Forlizzi, J. (2008). Robots in organizations. The Role of Workflow, Social, and Environmental Factors in Human-Robot Interaction. In T. Fong (ed.), *Proceedings of the 3rd ACMIEEE international conference on Human robot interaction* (S. 287–294). New York, NY: ACM.
- Mutlu, B., Roy, N. & Šabanović, S. (2016). Cognitive Human–Robot Interaction. In B. Siciliano & O. Khatib (Eds.), *Springer handbook of robotics* (2nd ed., S. 1907–1934). Berlin: Springer.
- Nagel, T. (1974). What Is It Like to Be a Bat? *The Philosophical Review*, 83(4), 435–450. <https://doi.org/10.2307/2183914>
- Nam, C. S., Eskander, E. & Choo, S. (2021). Neural dynamics of trust in human-robot interaction. In C. S. Nam & J. B. Lyons (Eds.), *Trust in Human-Robot Interaction* (S. 477–489). Elsevier Academic Press. <https://doi.org/10.1016/B978-0-12-819472-0.00020-4>
- Nam, C. S. & Lyons, J. B. (Eds.). (2021). *Trust in Human-Robot Interaction*: Elsevier Academic Press.
- Nass, C. I. & Brave, S. (2007). *Wired for speech. How voice activates and advances the human-computer relationship*. Cambridge Mass: MIT.
- Nass, C. I., Fogg, B. J. & Moon, Y. (1996). Can computers be teammates? *International Journal of Human-Computer Studies*, 45(6), 669–678. <https://doi.org/10.1006/ijhc.1996.0073>
- Nass, C. I. & Moon, Y. (2000). Machines and Mindlessness: Social Responses to Computers. *Journal of Social Issues*, 56(1), 81–103.
- Nerurkar, M., Wadepul, C. & Wieglerling, K. (2019). Metaphorik in der Technikethik: Ein Kommentar anlässlich der Big Data-Stellungnahme des Deutschen Ethikrats. In A. Friedrich, P. Gehring, C. Hubig, A. Kaminski & A. Nordmann (Hrsg.), *Steuern und Regeln* (S. 271–274). Nomos Verlagsgesellschaft mbH & Co. KG.
- Newman, G. E. (2018). Bringing Narratives to Life: Animism, Totems, and Intangible Value. *Journal of the Association for Consumer Research*, 3(4), 514–526. <https://doi.org/10.1086/699205>
- Neyer, F. J., Felber, J. & Gebhardt, C. (2012). Entwicklung und Validierung einer Kurzskala zur Erfassung von Technikbereitschaft. *Diagnostica*, 58(2), 87–99. <https://doi.org/10.1026/0012-1924/a000067>
- Nijssen, S. R. R., Müller, B. C. N., van Baaren, R. B. & Paulus, M. (2019). Saving the Robot or the Human? Robots Who Feel Deserve Moral Care. *Social Cognition*, 37(1), 41-56. <https://doi.org/10.1521/soco.2019.37.1.41>
- Nomura, T., Kanda, T., Suzuki, T. & Kato, K. (2004). Psychology in human-robot communication: an attempt through investigation of negative attitudes and anxiety toward

- robots. In *RO-MAN 2004. 13th IEEE International Workshop on Robot and Human Interactive Communication* (S. 35–40). <https://doi.org/10.1109/RO-MAN.2004.1374726>
- Nyholm, S. & Smids, J. (2020). Can a Robot Be a Good Colleague? *Science and Engineering Ethics*, 26(4), 2169–2188. <https://doi.org/10.1007/s11948-019-00172-6>
- O’Neill, O. (2018). Linking Trust to Trustworthiness. *International Journal of Philosophical Studies*, 26(2), 293–300. <https://doi.org/10.1080/09672559.2018.1454637>
- Obrist, M., Reitberger, W., Wurhofer, D., Förster, F. & Tscheligi, M. (2011). User Experience Research in the Semiconductor Factory: A Contradiction? In P. Campos, N. Graham, J. Jorge, N. Nunes, P. Palanque & M. Winckler (Eds.), *Human-computer interaction - INTERACT 2011. Proceedings of the 13th IFIP TC 13 international conference, Lisbon, Portugal, September 5 - 9, 2011, Part IV* (vol. 6949, S. 144–151). Berlin: Springer. https://doi.org/10.1007/978-3-642-23768-3_12
- Oistad, B. C., Sembroski, C. E., Gates, K. A., Krupp, M. M., Fraune, M. R. & Šabanović, S. (2016). Colleague or Tool? Interactivity Increases Positive Perceptions of and Willingness to Interact with a Robotic Co-worker. In A. Agah, A. M. Howard, M. A. Salichs, H. He & J.-J. Cabibihan (Eds.), *Social Robotics. Proceedings of the 8th International Conference, ICSR 2016 Kansas City, MO, USA, November 1–3, 2016* (S. 774–785). Cham: Springer International Publishing.
- Oliveira, R., Arriaga, P., Alves-Oliveira, P., Correia, F., Petisca, S. & Paiva, A. (2018). Friends or Foes? In T. Kanda, S. Šabanović, G. Hoffman & A. Tapus (eds.), *Proceedings of the 2018 ACM/IEEE International Conference on Human-Robot Interaction - HRI '18* (S. 279–288). New York, NY, USA: ACM Press.
- Onnasch, L., Jürgensohn, T., Remmers, P. & Asmuth, C. (2019). *Ethische und soziologische Aspekte der Mensch-Roboter-Interaktion*. baua: Bericht (Bundesanstalt für Arbeitsschutz und Arbeitsmedizin (BAuA), Hrsg.). Dortmund/Berlin/Dresden. <https://doi.org/10.21934/BAUA:BERICHT20190128>
- Onnasch, L., Maier, X. & Jürgensohn, T. (2016). *Mensch-Roboter-Interaktion - Eine Taxonomie für alle Anwendungsfälle*. baua: fokus. Dortmund: Bundesanstalt für Arbeitsschutz und Arbeitsmedizin. <https://doi.org/10.21934/baua:fokus20160630>
- Onnasch, L. & Roesler, E. (2019). Anthropomorphizing Robots: The Effect of Framing in Human-Robot Collaboration. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* (Bd. 63, S. 1311–1315). <https://doi.org/10.1177/1071181319631209>
- Onnasch, L. & Roesler, E. (2021). A Taxonomy to Structure and Analyze Human–Robot Interaction. *International Journal of Social Robotics*, 13(4), 833–849. Online-Vorabveröffentlichung. <https://doi.org/10.1007/s12369-020-00666-5>
- Osmont, A., Cassotti, M., Agogué, M., Houdé, O. & Moutier, S. (2015). Does ambiguity aversion influence the framing effect during decision making? *Psychonomic Bulletin & Review*, 22(2), 572–577. <https://doi.org/10.3758/s13423-014-0688-0>

- Osofsky, S. (2013). *Influence Of Task-role Mental Models On Human Interpretation Of Robot Motion Behavior*. Dissertation. University of Central Florida, Orlando. Zugriff am 19.04.2018. Verfügbar unter: <http://stars.library.ucf.edu/etd/2893>
- Osofsky, S., Phillips, E. K., Schuster, D. & Jentsch, F. G. (2013). A Picture is Worth a Thousand Mental Models. Evaluating Human Understanding of Robot Teammates. In *Proceedings of the Human Factors and Ergonomics Society 57th Annual Meeting* (S. 1298–1302). <https://doi.org/10.1177/1541931213571287>
- Ötting, S. K., Masjutin, L., Steil, J. J. & Maier, G. W. (2020). Let's Work Together: A Meta-Analysis on Robot Design Features That Enable Successful Human-Robot Interaction at Work. *Human Factors: the Journal of the Human Factors and Ergonomics Society*, 18720820966433. Online-Vorabveröffentlichung. <https://doi.org/10.1177/0018720820966433>
- Oubari, A., Pischke, D., Jenny, M., Meißner, A. & Trübswetter, A. (2018). Mensch-Roboter-Kollaboration in der Produktion. *ZWF Zeitschrift für wirtschaftlichen Fabrikbetrieb*, 113(9), 560–564. <https://doi.org/10.3139/104.111971>
- Paepcke, S. & Takayama, L. (2010). Judging a bot by its cover: An experiment on expectation setting for personal robots. In *5th ACM/IEEE International Conference on Human-Robot Interaction (HRI), 2010. March 2 - 5 2010, Osaka, Japan* (S. 45–52). Piscataway, NJ: IEEE.
- Parasuraman, R. & Riley, V. (1997). Humans and Automation: Use, Misuse, Disuse, Abuse. *Human Factors*, 39(2), 230–253. <https://doi.org/10.1518/00187209778543886>
- Persson, P., Laaksolahti, J. & Lonnqvist, P. (2000). *Anthropomorphism A Multi-Layered Phenomenon*. AAI Technical Report FS-00-04. Zugriff am 16.08.2020. Verfügbar unter: <https://www.aaai.org/Papers/Symposia/Fall/2000/FS-00-04/FS00-04-025.pdf>
- Peshkin, M. & Colgate, E. (1999a). Cobots. *Industrial Robot: An International Journal*, 26(5), 335–341.
- Peshkin, M. & Colgate, E. (1999b). *Cobots. United States Patent No. 5,952,796*.
- Phillips, E. K., Osofsky, S., Grove, J. & Jentsch, F. G. (2011). From Tools to Teammates. Toward the Development of Appropriate Mental Models for Intelligent Robots. In *Proceedings of the Human Factors and Ergonomics Society 55th Annual Meeting* (S. 1491–1495).
- Phillips, E. K., Schaefer, K. E., Billings, D. R., Jentsch, F. G. & Hancock, P. A. (2016). Human-Animal Teams as an Analog for Future Human-Robot Teams: Influencing Design and Fostering Trust. *Journal of Human-Robot Interaction*, 5(1), 100–125. <https://doi.org/10.5898/JHRI.5.1.Phillips>
- Phillips, E. K., Zhao, X., Ullman, D. & Malle, B. F. (2018). What is Human-like? Decomposing Robots' Human-like Appearance Using the Anthropomorphic roBOT (ABOT) Database. In T. Kanda, S. Šabanović, G. Hoffman & A. Tapus (eds.),

- Proceedings of the 2018 ACM/IEEE International Conference on Human-Robot Interaction - HRI '18* (S. 105–113). New York, NY, USA: ACM Press.
- Pöhler, G., Heine, T. & Deml, B. (2016). Itemanalyse und Faktorstruktur eines Fragebogens zur Messung von Vertrauen im Umgang mit automatischen Systemen. *Zeitschrift für Arbeitswissenschaft*, 70(3), 151–160. <https://doi.org/10.1007/s41449-016-0024-9>
- Powell, H. & Michael, J. (2019). Feeling committed to a robot: why, what, when and how? *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, 374, Aufsatz 20180039. <https://doi.org/10.1098/rstb.2018.0039>
- Powers, A. & Kiesler, S. (2006). The Advisor Robot: Tracing People's Mental Model from a Robot's Physical Attributes. In *Proceedings of the 2006 ACM Conference on Human-Robot Interaction* (S. 218–225).
- Preim, B. & Dachsel, R. (Hrsg.). (2010). *Interaktive Systeme* (eXamen.press, 2. Auflage). Berlin: Springer.
- Prill, Y. (2017). Speeder und Faker in Online-Befragungen. In A. Theobald (Hrsg.), *Praxis Online-Marktforschung* (S. 228–239). Wiesbaden: Springer Fachmedien Wiesbaden.
- Raab, G., Unger, A. & Unger, F. (2010). *Marktpsychologie. Grundlagen und Anwendung* (Gabler-Lehrbuch, 3., überarb. Aufl.). Wiesbaden: Gabler.
- Ragni, M. & Knauff, M. (2013). A theory and a computational model of spatial reasoning with preferred mental models. *Psychological review*, 120(3), 561–588. <https://doi.org/10.1037/a0032460>
- Rammstedt, B. & Danner, D. (2017). Die Facettenstruktur des Big Five Inventory (BFI). *Diagnostica*, 63(1), 70–84. <https://doi.org/10.1026/0012-1924/a000161>
- Ranz, F., Komenda, T., Reisinger, G., Hold, P., Hummel, V. & Sihm, W. (2018). A Morphology of Human Robot Collaboration Systems for Industrial Assembly. *Procedia CIRP*, 72, 99–104. <https://doi.org/10.1016/j.procir.2018.03.011>
- Reeves, B. & Nass, C. I. (1998). *The media equation. How people treat computers, television, and new media like real people and places* (1. paperback ed.). Stanford, Calif.: CSLI Publ.
- Rehm, M. & Krogsgager, A. (2013). Negative affect in human robot interaction — Impoliteness in unexpected encounters with robots. In *IEEE RO-MAN, 2013. The 22nd IEEE International Symposium on Robot and Human Interactive Communication; 26 - 29 Aug. 2013, Gyeongju, Korea* (S. 45–50). Piscataway, NJ: IEEE.
- Reiss, M. & Neumann, O. (2012). Konkurrierende Partner, aber kooperierende Wettbewerber. *Gestaltungsansätze für Coopetition im Ideen- und Innovationsmanagement. wissenschaftsmanagement*, (5), 48–51.
- Remmers, P. (2019). The Ethical Significance of Human Likeness in Robotics and AI. *Ethics in Progress*, 10(2), 52–67. <https://doi.org/10.14746/eip.2019.2.6>

- Remmers, P. (2020). Ethische Perspektiven der Mensch-Roboter-Kollaboration. In H.-J. Buxbaum (Hrsg.), *Mensch-Roboter-Kollaboration* (S. 55–68). Wiesbaden: Springer Fachmedien Wiesbaden. https://doi.org/10.1007/978-3-658-28307-0_4
- Richards, N. M. & Smart, W. D. (2016). How should the law think about robots? In R. Calo, M. A. Froomkin & I. Kerr (Eds.), *Robot Law* (S. 3–22). Cheltenham, UK: Edward Elgar Publishing.
- Richert, A. (2018). Socializing with Robots. In K. North, R. Maier & O. Haas (Eds.), *Knowledge Management in Digital Change. New Findings and Practical Cases* (97–110). Cham, Switzerland: Springer.
- Richert, A., Müller, S. L., Schröder, S. & Jeschke, S. (2017). Anthropomorphism in social robotics. Empirical results on human–robot interaction in hybrid production workplaces. *AI & SOCIETY*, 1(1), 71–80. <https://doi.org/10.1007/s00146-017-0756-x>
- Richert, A., Shehadeh, M. A., Müller, S. L., Schröder, S. & Jeschke, S. (2016). Robotic Workmates – Hybrid Human-Robot-Teams in the Industry 4.0. In *11th International Conference on e-Learning* (S. 127–131). Kuala Lumpur, Malaysia.
- Riek, L. D. (2012). Wizard of Oz Studies in HRI: A Systematic Review and New Reporting Guidelines. *Journal of Human-Robot Interaction*, 1(1), 119–136. <https://doi.org/10.5898/JHRI.1.1.Riek>
- Riek, L. D. & Howard, D. (2014). A Code of Ethics for the Human-Robot Interaction Profession. In *Proceedings of the WeRobot 2014*.
- Rizzolatti, G. & Craighero, L. (2004). The mirror-neuron system. *Annual Review of Neuroscience*, 27, 169–192. <https://doi.org/10.1146/annurev.neuro.27.070203.144230>
- Robla-Gomez, S., Becerra, V. M., Llata, J. R., Gonzalez-Sarabia, E., Torre-Ferrero, C. & Perez-Oria, J. (2017). Working Together: A Review on Safe Human-Robot Collaboration in Industrial Environments. *IEEE Access*, 5, 26754–26773. <https://doi.org/10.1109/ACCESS.2017.2773127>
- Roesler, E. & Onnasch, L. (2020a). The Effect of Anthropomorphism and Failure Comprehensibility on Human-Robot Trust. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 64(1), 107–111. <https://doi.org/10.1177/1071181320641028>
- Roesler, E. & Onnasch, L. (2020b). Teammitglied oder Werkzeug – Der Einfluss anthropomorpher Gestaltung in der Mensch-Roboter-Interaktion. In H.-J. Buxbaum (Hrsg.), *Mensch-Roboter-Kollaboration* (S. 163–175). Wiesbaden: Springer Fachmedien Wiesbaden. https://doi.org/10.1007/978-3-658-28307-0_11
- Rohrmann, B. (1978). Empirische Studien zur Entwicklung von Antwortskalen für die sozialwissenschaftliche Forschung. *Zeitschrift für Sozialpsychologie*, 9, 222–245.
- Rosenthal-von der Pütten, A. M., Schulte, F. P., Eimler, S. C., Sobieraj, S., Hoffmann, L., Maderwald, S. et al. (2014). Investigations on empathy towards humans and robots using fMRI. *Computers in Human Behavior*, 33, 201–212. <https://doi.org/10.1016/j.chb.2014.01.004>

- Rothbart, M. & Park, B. (1986). On the confirmability and disconfirmability of trait concepts. *Journal of Personality and Social Psychology*, 50(1), 131–142.
<https://doi.org/10.1037/0022-3514.50.1.131>
- Rotter, J. B. (1967). A new scale for the measurement of interpersonal trust. *Journal of Personality*, 35(4), 651–665. <https://doi.org/10.1111/j.1467-6494.1967.tb01454.x>
- Rouse, W. B., Cannon-Bowers, J. A. & Salas, E. (1992). The role of mental models in team performance in complex systems. *IEEE Transactions on Systems, Man, and Cybernetics*, 22(6), 1296–1308. <https://doi.org/10.1109/21.199457>
- Rouse, W. B. & Morris, N. M. (1986). On looking into the Black Box: Prospects and Limits in the Search for Mental Models. *Psychological Bulletin*, 100(3), 349–363.
- Royall, R. M. (1986). The Effect of Sample Size on the Meaning of Significance Tests. *The American Statistician*, 40(4), 313–315.
- Rueben, M., Bernieri, F. J., Grimm, C. M. & Smart, W. D. (2017). Framing Effects on Privacy Concerns about a Home Telepresence Robot. In B. Mutlu (ed.), *Proceedings of the 2017 ACM/IEEE International Conference on Human-Robot Interaction* (S. 435–444). New York, NY, USA: ACM Press.
- Ruxton, G. D. & Neuhäuser, M. (2010). When should we use one-tailed hypothesis testing? *Methods in Ecology and Evolution*, 1(2), 114–117.
<https://doi.org/10.1111/j.2041-210X.2010.00014.x>
- Šabanović, S., Reeder, S. & Kechavarzi, B. (2014). Designing Robots in the Wild: In situ Prototype Evaluation for a Break Management Robot. *Journal of Human-Robot Interaction*, 3(1), 70–88. <https://doi.org/10.5898/JHRI.3.1.Sabanovic>
- Salvini, P., Laschi, C. & Dario, P. (2010). Design for Acceptability. Improving Robots' Coexistence in Human Society. *International Journal of Social Robotics*, 2(4), 451–460. <https://doi.org/10.1007/s12369-010-0079-2>
- Samuel, J. L. (2019). Company from the Uncanny Valley: A Psychological Perspective on Social Robots, Anthropomorphism and the Introduction of Robots to Society. *Ethics in Progress*, 10(2), 8–26. <https://doi.org/10.14746/eip.2019.2.2>
- Sanders, T. L., Kaplan, A. D., Koch, R., Schwartz, M. & Hancock, P. A. (2019). The Relationship Between Trust and Use Choice in Human-Robot Interaction. *Human Factors*, 61(4), 614–626. <https://doi.org/10.1177/0018720818816838>
- Sanders, T. L., MacArthur, K., Volante, W., Hancock, G., MacGillivray, T., Shugars, W. et al. (2017). Trust and Prior Experience in Human-Robot Interaction. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 61(1), 1809–1813.
<https://doi.org/10.1177/1541931213601934>
- Sanders, T. L., Wixon, T., Schaefer, K. E., Chen, J. Y. C. & Hancock, P. A. (2014). The influence of modality and transparency on trust in human-robot interaction. In *2014 IEEE International Inter-Disciplinary Conference on Cognitive Methods in Situation Awareness and Decision Support (CogSIMA 2014)*. San Antonio, Texas, USA, March 3-6 2014 (S. 156–159). Piscataway, NJ: IEEE.

- Sandoval, E. B., Brandstetter, J., Obaid, M. & Bartneck, C. (2016). Reciprocity in Human-Robot Interaction. A Quantitative Approach Through the Prisoner's Dilemma and the Ultimatum Game. *International Journal of Social Robotics*, 8(2), 303–317. <https://doi.org/10.1007/s12369-015-0323-x>
- Sauppé, A. & Mutlu, B. (2015). The Social Impact of a Robot Co-Worker in Industrial Settings. In J. Kim (ed.), *CHI 2015 crossings. Proceedings of the 33rd Annual CHI Conference on Human Factors in Computing Systems; April 18 - 23, 2015, Seoul, Republic of Korea* (S. 3613–3622). New York, NY, USA: ACM.
- Savela, N., Turja, T. & Oksanen, A. (2018). Social Acceptance of Robots in Different Occupational Fields: A Systematic Literature Review. *International Journal of Social Robotics*, 10(4), 493–502. <https://doi.org/10.1007/s12369-017-0452-5>
- Schaefer, K. E. (2013). *The Perception And Measurement Of Human-robot Trust*. Dissertation. University of Central Florida.
- Schaefer, K. E. (2016). Measuring Trust in Human Robot Interactions: Development of the “Trust Perception Scale-HRI”. In R. Mittu, D. Sofge, A. Wagner & W. F. Lawless (Eds.), *Robust intelligence and trust in autonomous systems* (vol. 1, S. 191–218). New York: Springer. https://doi.org/10.1007/978-1-4899-7668-0_10
- Schaefer, K. E., Chen, J. Y. C., Szalma, J. L. & Hancock, P. A. (2016). A Meta-Analysis of Factors Influencing the Development of Trust in Automation: Implications for Understanding Autonomy in Future Systems. *Human Factors*, 58(3), 377–400. <https://doi.org/10.1177/0018720816634228>
- Schaefer, K. E., Perelman, B. S., Gremillion, G. M., Marathe, A. R. & Metcalfe, J. S. (2021). A roadmap for developing team trust metrics for human-autonomy teams. In C. S. Nam & J. B. Lyons (Eds.), *Trust in Human-Robot Interaction* (S. 261–300). Elsevier Academic Press. <https://doi.org/10.1016/B978-0-12-819472-0.00012-5>
- Schellen, E., Pérez-Osorio, J. & Wykowska, A. (2018). Social Cognition in Human-Robot Interaction: Putting the ‘H’ back in ‘HRI’. *ERCIM News*, (114), 23–24.
- Schellen, E. & Wykowska, A. (2019). Intentional Mindset Toward Robots—Open Questions and Methodological Challenges. *Frontiers in Robotics and AI*, 5, Aufsatz 139. <https://doi.org/10.3389/frobt.2018.00139>
- Scheutz, M., DeLoach, S. A. & Adams, J. A. (2017). A Framework for Developing and Using Shared Mental Models in Human-Agent Teams. *Journal of Cognitive Engineering and Decision Making*, 11(3), 203–224. <https://doi.org/10.1177/1555343416682891>
- Schou, C., Andersen, R. S., Chrysostomou, D., Bøgh, S. & Madsen, O. (2018). Skill-based instruction of collaborative robots in industrial settings. *Robotics and Computer-Integrated Manufacturing*, 53, 72–80. <https://doi.org/10.1016/j.rcim.2018.03.008>

- Schou, C. & Madsen, O. (2017). A plug and produce framework for industrial collaborative robots. *International Journal of Advanced Robotic Systems*, 14(4), 1-10.
<https://doi.org/10.1177/1729881417717472>
- Schreier, M. (2014). Varianten qualitativer Inhaltsanalyse: Ein Wegweiser im Dickicht der Begrifflichkeiten. *FQS - Forum Qualitative Sozialforschung*, 15(1), Aufsatz 18.
<https://doi.org/10.17169/fqs-15.1.2043>
- Schröder, J. & Epley, N. (2016). Mistaking minds and machines: How speech affects dehumanization and anthropomorphism. *Journal of Experimental Psychology. General*, 145(11), 1427–1437. <https://doi.org/10.1037/xge0000214>
- Schuster, D., Ososky, S., Jentsch, F. G., Phillips, E. K., Lebiere, C. & Evans, W. A. (2011). A Research Approach to Shared Mental Models and Situation Assessment in Future Robot Teams. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 55(1), 456–460. <https://doi.org/10.1177/1071181311551094>
- Seel, N. M. (2001). Epistemology, situated cognition, and mental models: 'Like a bridge over troubled water'. *Instructional Science*, 29(4/5), 403–427.
<https://doi.org/10.1023/A:1011952010705>
- Seel, N. M. (2017). Model-based learning: a synthesis of theory and research. *Educational Technology Research and Development*, 65(4), 931–966.
<https://doi.org/10.1007/s11423-016-9507-9>
- Seel, N. M., Ifenthaler, D. & Pirnay-Dummer, P. (2009). Mental Models and Problem Solving: Technological Solutions for Measurement and Assessment of the Development of Expertise. In P. Blumschein, W. Hung, D. Jonasson & J. Strobel (Eds.), *Model-bases approaches to learning. Using systems models and stimulations to improve understanding and problem solving in complex domains* (Modeling and simulations for learning and instruction, vol. 4, S. 17–40). Rotterdam u.a.: Sense Publ.
- Seng, L. (2019). Maschinenethik und Künstliche Intelligenz. In O. Bendel (Hrsg.), *Handbuch Maschinenethik* (Springer Reference Geisteswissenschaften, 1st ed. 2019, S. 1–21). Wiesbaden: Springer Fachmedien. https://doi.org/10.1007/978-3-658-17484-2_13-1
- Shehadeh, M. A., Schröder, S., Richert, A. & Jeschke, S. (2017). Hybrid teams of industry 4.0. A work place considering robots as key players. In *2017 IEEE International Conference on Systems, Man, and Cybernetics (SMC). Banff Center, Banff, Canada, October 5-8, 2017* (S. 1208–1213). Piscataway, N.J: IEEE.
- Sheridan, T. B. (2016). Human-Robot Interaction. Status and Challenges. *Human Factors*, 58(4), 525–532. <https://doi.org/10.1177/0018720816644364>
- Siciliano, B. & Khatib, O. (Eds.). (2016). *Springer handbook of robotics* (2nd ed.). Berlin: Springer.
- Siegel, M., Breazeal, C. L. & Norton, M. I. (2009). Persuasive Robotics: The influence of robot gender on human behavior. In *2009 IEEE/RSJ International Conference on*

- Intelligent Robots and Systems. St Louis, Missouri, 10-15 October 2009* (S. 2563–2568). Piscataway, NJ: IEEE.
- Skevi, A., Szigeti, H., Perini, S., Oliveira, M., Taisch, M. & Kiritsis, D. (2014). Current Skills Gap in Manufacturing: Towards a New Skills Framework for Factories of the Future. In B. Grabot, B. Vallespir, S. Gomes, A. Bouras & D. Kiritsis (eds.), *Advances in Production Management Systems. Innovative and Knowledge-Based Production Management in a Global-Local World* (vol. 438, S. 175–183). Berlin, Heidelberg: Springer Berlin Heidelberg.
- Smids, J., Nyholm, S. & Berkers, H. (2020). Robots in the Workplace: a Threat to—or Opportunity for—Meaningful Work? *Philosophy & Technology*, 33(3), 503–522. <https://doi.org/10.1007/s13347-019-00377-4>
- Smith, M. A. B. & Wiese, E. (2016). Look at Me Now: Investigating Delayed Disengagement for Ambiguous Human-Robot Stimuli. In A. Agah, A. M. Howard, M. A. Sallichs, H. He & J.-J. Cabibihan (Eds.), *Social Robotics. Proceedings of the 8th International Conference, ICSR 2016 Kansas City, MO, USA, November 1–3, 2016* (vol. 9979, S. 950–960). Cham: Springer International Publishing. https://doi.org/10.1007/978-3-319-47437-3_93
- Solo, B. (2013). *Math Gear(s)*, MakerBot's Thingiverse. Zugriff am 17.01.2022. Verfügbar unter: <https://www.thingiverse.com/thing:185912>
- Spain, R. D., Bustamante, E. A. & Bliss, J. P. (2008). Towards an empirically developed scale for system trust: take two. In *Proceedings of the Human Factors and Ergonomics Society 52nd annual meeting. September 22 - 26, 2008, New York City, NY, USA* (S. 1335–1339). Santa Monica, Calif.: Human Factors and Ergonomics Soc.
- Stadler, S., Weiss, A., Mirmig, N. & Tscheligi, M. (2013). Anthropomorphism in the factory - a paradigm change? In *HRI 2013. Proceedings of the 8th ACM/IEEE International Conference on Human-Robot Interaction* (S. 231–232). Piscataway, NJ: IEEE.
- Staggers, N. & Norcio, A. F. (1993). Mental models: concepts for human-computer interaction research. *Int. J. Man-Machine Studies (International Journal of Man-Machine Studies)*, 38(4), 587–605.
- Statistisches Bundesamt Deutschland. (2020, 15. Dezember). *Beschäftigte und Umsatz der Betriebe im Verarbeitenden Gewerbe: Deutschland, Jahre, Wirtschaftszweige (WZ2008 2-/3-/4-Steller)*, Statistisches Bundesamt Deutschland. Zugriff am 15.12.2020. Verfügbar unter: <https://www-genesis.destatis.de/genesis/online?operation=previous&levelindex=1&step=1&titel=Ergebnis&levelid=1608025250574&acceptcookies=false#abreadcrumb>
- Steil, J. J. & Maier, G. W. (2017a). Kollaborative Roboter: universale Werkzeuge in der digitalisierten und vernetzten Arbeitswelt. In G. W. Maier, G. Engels & E. Steffen (Hrsg.), *Handbuch Gestaltung digitaler und vernetzter Arbeitswelten* (Springer Reference Psychologie, Living reference work, continuously updated edition, Bd. 57, S. 1–24). Berlin: Springer.

- Steil, J. J. & Maier, G. W. (2017b). Robots in the Digitalized Workplace. In G. Hertel (Ed.), *The Wiley Blackwell handbook of the psychology of the Internet at work* (Wiley Blackwell handbooks in organizational psychology, vol. 19, S. 401–422). Hoboken, NJ: John Wiley & Sons Ltd. <https://doi.org/10.1002/9781119256151.ch18>
- Stenzel, A., Chinellato, E., Bou, M. A. T., del Pobil, Á. P., Lappe, M. & Liepelt, R. (2012). When humanoid robots become human-like interaction partners: corepresentation of robotic actions. *Journal of Experimental Psychology. Human Perception and Performance*, 38(5), 1073–1077. <https://doi.org/10.1037/a0029493>
- Stenzel, A., Chinellato, E., del Pobil, Á. P., Lappe, M. & Liepelt, R. (2012). How deeply do we include Robotic Agents in the Self? *International Journal of Humanoid Robotics*, 10(1), 1–13.
- Stocké, V. (2002). *Framing und Rationalität. Die Bedeutung der Informationsdarstellung für das Entscheidungsverhalten* (Scientia Nova). Zugl.: Mannheim, Univ., Dissertationsschrift, 2000. München: De Gruyter Oldenbourg. <https://doi.org/10.1524/9783486833263>
- Strohkorb Sebo, S., Traeger, M., Jung, M. F. & Scassellati, B. (2018). The Ripple Effects of Vulnerability. In T. Kanda, S. Šabanović, G. Hoffman & A. Tapus (eds.), *Proceedings of the 2018 ACM/IEEE International Conference on Human-Robot Interaction - HRI '18* (S. 178–186). New York, NY, USA: ACM Press.
- Sumpf, P. (2019). *System Trust. Researching the Architecture of Trust in Systems*. Springer Fachmedien Wiesbaden.
- Syrdal, D. S., Dautenhahn, K., Koay, K. L., Walters, M. L. & Otero, N. R. (2010). Exploring human mental models of robots through explicitation interviews. In *IEEE RO-MAN, 2010. 19th IEEE International Symposium on Robot and Human Interactive Communication; September 13-15 2010, Viareggio, Italy* (S. 638–645). Piscataway, NJ: IEEE.
- Der Tagesspiegel. (2019). *Toyota feuert die Roboter*, Der Tagesspiegel. Zugriff am 06.01.2019. Verfügbar unter: <https://www.tagesspiegel.de/themen/reportage/kuenstliche-intelligenz-toyota-feuert-die-roboter/23821418.html>
- Talone, A. B., Phillips, E. K., Ososky, S. & Jentsch, F. G. (2016). An Evaluation of Human Mental Models of Tactical Robot Movement. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 59(1), 1558–1562. <https://doi.org/10.1177/1541931215591337>
- Tanner, A., Burkhard, R. & Schulze, H. (2019). Soziale Roboter – Erfolgsfaktoren für die Umsetzung ihrer Potenziale. Ergebnisse einer Fallstudie in der Schweiz. In Gesellschaft für Arbeitswissenschaft (Hrsg.), *GfA-Frühjahrskongress 2019. Arbeit interdisziplinär analysieren - bewerten - gestalten*.
- Tauer, J. M. & Harackiewicz, J. M. (2004). The effects of cooperation and competition on intrinsic motivation and performance. *Journal of Personality and Social Psychology*, 86(6), 849–861. <https://doi.org/10.1037/0022-3514.86.6.849>

- Tausch, N., Kenworthy, J. B. & Hewstone, M. (2007). The confirmability and disconfirmability of trait concepts revisited: does content matter? *Journal of Personality and Social Psychology*, 92(3), 542–556. <https://doi.org/10.1037/0022-3514.92.3.542>
- Tay, B., Jung, Y. & Park, T. (2014). When stereotypes meet robots: The double-edge sword of robot gender and personality in human–robot interaction. *Computers in Human Behavior*, 38, 75–84. <https://doi.org/10.1016/j.chb.2014.05.014>
- Theobald, A. (Hrsg.). (2017). *Praxis Online-Marktforschung*. Wiesbaden: Springer Fachmedien Wiesbaden. <https://doi.org/10.1007/978-3-658-10203-6>
- Tinsley, H. E. A. & Brown, S. D. (2000). *Handbook of applied multivariate statistics and mathematical modeling*. San Diego, Calif.: Academic Press. <https://doi.org/10.1016/B978-0-12-691360-6.X5000-9>
- Trafimow, D. & Sheeran, P. (2004). A theory about the translation of cognition into affect and behavior. In G. Haddock & G. R. Maio (Eds.), *Contemporary Perspectives on the Psychology of Attitudes* (S. 57–76). Psychology Press.
- Trübswetter, A., Meißner, A., Weber, M.-A., Klues, J. & Stowasser, S. (2018). Kollaborierende Roboter in der Produktion - Akzeptanz durch die Beschäftigten. *Betriebspraxis & Arbeitsforschung*, (233), 24–27.
- Turja, T. & Oksanen, A. (2019). Robot Acceptance at Work: A Multilevel Analysis Based on 27 EU Countries. *International Journal of Social Robotics*, 11(4), 679–689. <https://doi.org/10.1007/s12369-019-00526-x>
- Turkle, S. (2002). Sociable Technologies: Enhancing Human Performance when the computer is not a tool but a companion. In M. C. Roco & W. Sims Bainbridge (Eds.), *Converging Technologies for Improving Human Performance. Nanotechnology, biotechnology, information technology and cognitive science* (150-158).
- Turkle, S. (2005). *The second self. Computers and the human spirit* (20th anniversary ed., 1st MIT Press ed.). Cambridge Mass.: MIT Press.
- Turkle, S. (Ed.). (2007). *Evocative objects. Things we think with* (1. paperback edition). Cambridge, Mass.: The MIT Press.
- Turkle, S. (2010). In good company? On the threshold of robotic companions. In Y. Wilks (Ed.), *Close engagements with artificial companions. Key social, psychological, ethical and design issues* (Natural language processing, vol. 8). Philadelphia, PA: John Benjamins.
- Turkle, S. (2011). *Alone together. Why we expect more from technology and less from each other*. New York: Basic Books.
- Tversky, A. & Kahneman, D. (1981). The framing of decisions and the psychology of choice. *Science (New York, N.Y.)*, 211(4481), 453–458. <https://doi.org/10.1126/science.7455683>
- Tversky, A. & Kahneman, D. (1986). Rational Choice and the Framing of Decisions. *Journal of Business*, 59(4), 5251–5278.

- Ullman, D. & Malle, B. F. (2017). Human-Robot Trust: Just a Button Press Away. In B. Mutlu, M. Tscheligi, A. Weiss & J. E. Young (eds.), *Proceedings of the Companion of the 2017 ACM/IEEE International Conference on Human-Robot Interaction - HRI '17* (S. 309–310). New York, NY, USA: ACM Press.
- Ullman, D. & Malle, B. F. (2018). What Does it Mean to Trust a Robot? In T. Kanda, S. Šabanović, G. Hoffman & A. Tapus (eds.), *Proceedings of the 2018 ACM/IEEE International Conference on Human-Robot Interaction - HRI '18* (S. 263–264). New York, NY, USA: ACM Press.
- Ullman, D. & Malle, B. F. (2019). Measuring Gains and Losses in Human-Robot Trust: Evidence for Differentiable Components of Trust. In *Proceedings of the 14th ACM/IEEE International Conference on Human-Robot Interaction* (S. 618–619). IEEE Press.
- Ullrich, A., Vladova, G., Thim, C. & Gronau, N. (2015). Akzeptanz und Wandlungsfähigkeit im Zeichen der Industrie 4.0. *HMD Praxis der Wirtschaftsinformatik*, 52(5), 769–789. <https://doi.org/10.1365/s40702-015-0167-8>
- Vanderborght, B., Berte, J., Coppel, G. de, El Makrini, I., Elprama, S. A., Jewell, C. I. et al. (2017). Towards an Acceptable Socially Collaborative Robot for the Manufacturing Industry. In *ICRA workshop IC3 - Industry of the future: Collaborative, Connected, Cognitive: Novel approaches stemming from Factory of the Future & Industry 4.0 initiatives* (S. 1–2).
- Venkatesh, V. & Bala, H. (2008). Technology Acceptance Model 3 and a Research Agenda on Interventions. *Decision Sciences*, 39(2), 273–315. <https://doi.org/10.1111/j.1540-5915.2008.00192.x>
- Venkatesh, V. & Davis, F. D. (2000). A Theoretical Extension of the Technology Acceptance Model: Four Longitudinal Field Studies. *Management Science*, 46(2), 186–204. <https://doi.org/10.1287/mnsc.46.2.186.11926>
- Venkatesh, V., Morris, M. G., Davis, G. B. & Davis, F. D. (2003). User Acceptance of Information Technology: Toward a Unified View. *MIS Quarterly*, 27(3), 425–478. <https://doi.org/10.2307/30036540>
- Venkatesh, V., Thong, J. Y. L. & Xu, X. (2012). Consumer Acceptance and Use of Information Technology: Extending the Unified Theory of Acceptance and Use of Technology. *MIS Quarterly*, 36(1), 157–178. <https://doi.org/10.2307/41410412>
- Vincent, J., Taipale, S., Sapio, B., Lugano, G. & Fortunati, L. (2015). *Social Robots from a Human Perspective*. Cham: Springer International Publishing. <https://doi.org/10.1007/978-3-319-15672-9>
- Wadephul, C. (2020). (Bio- und anthropomorphe) Metaphorik in der Rede über KI und Robotik - Vermeidbares Missverständnis oder notwendiges Übel? In N. Hemlein (Hrsg.), *KI und interkulturelle Robotik – Online Tagung & Workshop 31.11. – 01.12.2020* (S. 28–31).

- Wainer, H. (1991). Adjusting for differential base rates: Lord's paradox again. *Psychological Bulletin*, 109(1), 147–151. <https://doi.org/10.1037/0033-2909.109.1.147>
- Wang, Y., Ma, H.-S., Yang, J.-H. & Wang, K.-S. (2017). Industry 4.0: a way from mass customization to mass personalization production. *Advances in Manufacturing*, 5(4), 311–320. <https://doi.org/10.1007/s40436-017-0204-7>
- Washburn, A., Adeleye, A., An, T. & Riek, L. D. (2020). Robot Errors in Proximate HRI: How Functionality Framing Affects Perceived Reliability and Trust. *ACM Transactions on Human-Robot Interaction*, 9(3), 1–21. <https://doi.org/10.1145/3380783>
- Waytz, A., Cacioppo, J. T. & Epley, N. (2010). Who Sees Human? The Stability and Importance of Individual Differences in Anthropomorphism. *Perspectives on Psychological Science: a Journal of the Association for Psychological Science*, 5(3), 219–232. <https://doi.org/10.1177/1745691610369336>
- Waytz, A., Heafner, J. & Epley, N. (2014). The mind in the machine: Anthropomorphism increases trust in an autonomous vehicle. *Journal of Experimental Social Psychology*, 52, 113–117. <https://doi.org/10.1016/j.jesp.2014.01.005>
- Waytz, A., Morewedge, C. K., Epley, N., Monteleone, G., Gao, J.-H. & Cacioppo, J. T. (2010). Making sense by making sentient: effectance motivation increases anthropomorphism. *Journal of Personality and Social Psychology*, 99(3), 410–435. <https://doi.org/10.1037/a0020240>
- Waytz, A. & Norton, M. I. (2014). Botsourcing and outsourcing: Robot, British, Chinese, and German workers are for thinking--not feeling--jobs. *Emotion*, 14(2), 434–444. <https://doi.org/10.1037/a0036054>
- Wehling, E. (2016). *Politisches Framing. Wie eine Nation sich ihr Denken einredet - und daraus Politik macht* (Edition medienpraxis, Bd. 14). Köln: Herbert von Halem Verlag.
- Weis, P. P. & Wiese, E. (2020). Know Your Cognitive Environment! Mental Models as Crucial Determinant of Offloading Preferences. *Human Factors*. Online-Vorabveröffentlichung. <https://doi.org/10.1177/0018720820956861>
- Weiss, A., Bernhaupt, R., Tscheligi, M., Wollherr, D., Kühnlenz, K. & Buss, M. (2008). A Methodological Variation for Acceptance Evaluation of Human-Robot Interaction in Public Places. In *Proceedings of the 17th IEEE International Symposium on Robot and Human Interactive Communication* (S. 713–718). Munich, Germany.
- Weiss, A. & Huber, A. (2016). User Experience of a Smart Factory Robot: Assembly Line Workers Demand Adaptive Robots. *5th International Symposium on New Frontiers in Human-Robot Interaction 2016*.
- Weiss, A., Huber, A., Minichberger, J. & Ikeda, M. (2016). First Application of Robot Teaching in an Existing Industry 4.0 Environment. Does It Really Work? *Societies*, 6(3), 1–21. <https://doi.org/10.3390/soc6030020>
- Wen, J., Stewart, A., Billingham, M., Dey, A., Tossell, C. C. & Finomore, V. S. (2018). He who hesitates is lost (...in thoughts over a robot). In *Proceedings of the*

- Technology, Mind, and Society on ZZZ - TechMindSociety '18* (S. 1–6). New York, New York, USA: ACM Press.
- Wiese, E., Metta, G. & Wykowska, A. (2017). Robots As Intentional Agents. Using Neuroscientific Methods to Make Robots Appear More Social. *Frontiers in Psychology*, 8, Aufsatz 1663. <https://doi.org/10.3389/fpsyg.2017.01663>
- Wilcox, R., Nikolaidis, S. & Shah, J. A. (2013). Optimization of Temporal Dynamics for Adaptive Human-Robot Interaction in Assembly Manufacturing. In N. Roy (ed.), *Robotics. Science and systems VIII* (S. 441–448). Cambridge, Mass: MIT Press.
- Wilkinson, B. D., Fisk, A. D. & Rogers, W. A. (2007). Effects of Mental Model Quality on Collaborative System Performance. *Proceedings of the Human Factors and Ergonomics Society 51st Annual Meeting 2007*, 51(22), 1506–1510.
- Wilson, J. R. & Rutherford, A. (1989). Mental Models: Theory and Application in Human Factors. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 31(6), 617–634. <https://doi.org/10.1177/001872088903100601>
- Winter, K. (2013). *Soziale und sozialpsychische Determinanten des Gesundheitsverhaltens. Eine theoriegeleitete Analyse am Beispiel von Überernährung und Tabakkonsum*. Dissertation. Freie Universität Berlin, Berlin.
- Wischniewski, S., Rosen, P. H. & Kirchhoff, B. (2019). Stand der Technik und zukünftige Entwicklungen der Mensch-Roboter-Interaktion. In Gesellschaft für Arbeitswissenschaft (Hrsg.), *GfA-Frühjahrskongress 2019. Arbeit interdisziplinär analysieren - bewerten - gestalten*. (C.10.11).
- Wu, K., Zhao, Y., Zhu, Q., Tan, X. & Zheng, H. (2011). A meta-analysis of the impact of trust on technology acceptance model: Investigation of moderating influence of subject and context type. *International Journal of Information Management*, 31(6), 572–581. <https://doi.org/10.1016/j.ijinfomgt.2011.03.004>
- Wullenkord, R. (2017). *Messung und Veränderung von Einstellungen gegenüber Robotern – Untersuchung des Einflusses von imaginiertem Kontakt auf implizite und explizite Maße*. Dissertation. Bielefeld.
- Wullenkord, R. & Eyssel, F. (2014). Improving attitudes towards social robots using imagined contact. In *RO-MAN: the 23rd IEEE International Symposium on Robot and Human Interactive Communication* (S. 489–494).
- Wullenkord, R. & Eyssel, F. (2019). Imagine how to behave: the influence of imagined contact on human-robot interaction. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, (374), Aufsatz 1771. <https://doi.org/10.1098/rstb.2018.0038>
- Wullenkord, R., Fraune, M. R., Eyssel, F. & Šabanović, S. (2016). Getting in Touch: How Imagined, Actual, and Physical Contact Affect Evaluations of Robots. In *25th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)* (S. 980–985).

- Wurhofer, D., Meneweger, T., Fuchsberger, V. & Tscheligi, M. (2015). Deploying Robots in a Production Environment: A Study on Temporal Transitions of Workers' Experiences. In J. Abascal González, S. Barbosa, M. Fetter, T. Gross, P. Palanque & M. Winckler (Eds.), *Proceedings of the 15th IFIP TC 13 International Conference, Bamberg, Germany, September 14 - 18, 2015* (vol. 9298, S. 203–220). Cham: Springer. https://doi.org/10.1007/978-3-319-22698-9_14
- Wynne, K. T. & Lyons, J. B. (2018). An integrative model of autonomous agent teammate-likeness. *Theoretical Issues in Ergonomics Science*, 19(3), 353–374. <https://doi.org/10.1080/1463922X.2016.1260181>
- Wynne, K. T. & Lyons, J. B. (2019). Autonomous Agent Teammate-Likeness: Scale Development and Validation. In J. Y. C. Chen & G. Fragomeni (Eds.), *Virtual, Augmented and Mixed Reality. Proceedings of Multimodal interaction: 11th international conference, VAMR 2019: held as part of the 21st HCI international conference, HCII 2019 : Orlando, FL, USA, July 26-31, 2019* (vol. 11575, S. 199–213). Cham: Springer International Publishing. https://doi.org/10.1007/978-3-030-21565-1_13
- Xing, B. & Marwala, T. (2018). Introduction to Human Robot Interaction. In B. Xing & T. Marwala (Eds.), *Smart maintenance for human-robot interaction. An intelligent search algorithmic perspective* (Studies in Systems, Decision and Control, vol. 129, S. 3–19). Cham: Springer-Verlag. https://doi.org/10.1007/978-3-319-67480-3_1
- Yagoda, R. E. & Gillan, D. J. (2012). You Want Me to Trust a ROBOT? The Development of a Human–Robot Interaction Trust Scale. *International Journal of Social Robotics*, 4(3), 235–248. <https://doi.org/10.1007/s12369-012-0144-0>
- Yang, X. J., Unhelkar, V. V., Li, K. & Shah, J. A. (2017). Evaluating Effects of User Experience and System Transparency on Trust in Automation. In B. Mutlu (ed.), *Proceedings of the 2017 ACM/IEEE International Conference on Human-Robot Interaction* (S. 408–416). New York, NY, USA: ACM Press.
- Yogeeswaran, K., Zlotowski, J., Livingstone, M., Bartneck, C., Sumioka, H. & Ishiguro, H. (2016). The Interactive Effects of Robot Anthropomorphism and Robot Ability on Perceived Threat and Support for Robotics Research. *Journal of Human-Robot Interaction*, 5(2), 29–47. <https://doi.org/10.5898/JHRI.5.2.Yogeeswaran>
- You, S. & Robert, L. P. (2018). Human–Robot Similarity and Willingness to Work with a Robotic Co-worker. In T. Kanda, S. Šabanović, G. Hoffman & A. Tapus (eds.), *Proceedings of the 2018 ACM/IEEE International Conference on Human-Robot Interaction - HRI '18* (S. 251–260). New York, NY, USA: ACM Press.
- You, S. & Robert, L. P. (2019). Trusting Robots in Teams: Examining the Impacts of Trusting Robots on Team Performance and Satisfaction. Forthcoming. In *Proceedings of the 52th Hawaii International Conference on System Sciences (HICSS 2019)*.
- Zawieska, K., Duffy, B. R. & Spronska, A. (2012). Understanding anthropomorphisation in social robotics. *Pomiarowy automatyka Robotyka*, (11), 78–82.

- Ziegler, R. (2010). Ambiguität und Ambivalenz in der Psychologie. Begriffsverständnis und Begriffsverwendung. *Zeitschrift für Literaturwissenschaft und Linguistik*, 40(158), 125–171.
- Zijlstra, H. (2017). *Equality for All? Discursive Anthropomorphic Framing in Social Robotics*. Master Thesis. Utrecht University, Utrecht.
- Złotowski, J., Proudfoot, D., Yogeewaran, K. & Bartneck, C. (2015). Anthropomorphism: Opportunities and Challenges in Human–Robot Interaction. *International Journal of Social Robotics*, (7), 347–360.
- Złotowski, J., Strasser, E. & Bartneck, C. (2014). Dimensions of anthropomorphism. In G. Sagerer, M. Imai, T. Belpaeme & A. L. Thomaz (eds.), *HRI 2014. Proceedings of the 2014 ACM/IEEE International Conference on Human-Robot Interaction* (S. 66–73). New York, NY: ACM.
- Złotowski, J., Sumioka, H., Bartneck, C., Nishio, S. & Ishiguro, H. (2017). Understanding Anthropomorphism: Anthropomorphism is not a Reverse Process of Dehumanization. In A. Kheddar, E. Yoshida, S. S. Ge, K. Suzuki, J.-J. Cabibihan, F. Eyssel et al. (Eds.), *Social Robotics. 9th International Conference, ICSR 2017, Tsukuba, Japan, November 22-24, 2017, Proceedings* (Lecture Notes in Computer Science, vol. 10652, S. 618–627). Cham: Springer International Publishing. https://doi.org/10.1007/978-3-319-70022-9_61
- Złotowski, J., Sumioka, H., Eyssel, F., Nishio, S., Bartneck, C. & Ishiguro, H. (2018). Model of Dual Anthropomorphism: The Relationship Between the Media Equation Effect and Implicit Anthropomorphism. *International Journal of Social Robotics*, 34(4), 1–14. <https://doi.org/10.1007/s12369-018-0476-5>
- Złotowski, J., Sumioka, H., Nishio, S., Glas, D. F., Bartneck, C. & Ishiguro, H. (2015). Persistence of the uncanny valley: the influence of repeated interactions and a robot's attitude on its perception. *Frontiers in Psychology*, 6, Aufsatz 883. <https://doi.org/10.3389/fpsyg.2015.00883>

Anhang

A. Veröffentlichungshinweise

Parallel zur Anfertigung dieser Dissertation wurden einzelne Inhalte im Rahmen von Artikeln in Fachzeitschriften veröffentlicht. Tabelle 42 listet diese Publikationen auf und ordnet sie den entsprechenden Kapiteln der Dissertation zu.

Tabelle 42: Tabellarische Darstellung der in der Dissertation verwendeten eigenen Veröffentlichungen.

Veröffentlichung	Kapitel
Kopp, T., Schäfer, A. & Kinkel, S. (2020). Kollaborierende oder kollaborationsfähige Roboter? Welche Rolle spielt die Mensch-Roboter-Kollaboration in der Praxis? <i>Industrie 4.0 Management</i> , 2020(2), 19–23. https://doi.org/10.30844/I40M_20-2_S19-23	2.1.2; 2.2.3.2; 4.3
Kopp, T., Baumgartner, M. & Kinkel, S. (2021). Success factors for introducing industrial human-robot interaction in practice: an empirically driven framework. <i>The International Journal of Advanced Manufacturing Technology</i> , 112(3-4), 685–704. https://doi.org/10.1007/s00170-020-06398-0	4.2
Kopp, T., Baumgartner, M. & Kinkel, S. (2022). How Linguistic Framing Affects Factory Workers' Initial Trust in Collaborative Robots: The Interplay Between Anthropomorphism and Technological Replacement. <i>International Journal of Human-Computer Studies</i> , 158, Aufsatz 102730. https://doi.org/10.1016/j.ijhcs.2021.102730	3.4.7.4; 4.5; 6.1.1
Kopp, T. (2020). Die Wirkkraft der Sprache auf die Wahrnehmung neuartiger Technologien wie kollaborationsfähiger Roboter (Cobots) oder künstlicher Intelligenz (KI). In F. Nees, I. Stengel, V. G. Meister, T. Barton, F. Herrmann, C. Müller et al. (Hrsg.), <i>Angewandte Forschung in der Wirtschaftsinformatik 2020. Tagungsband zur 33. AKWI-Jahrestagung am 14.09.2020</i> (S. 41–52). Heide: mana Buch.	6.2

Veröffentlichung	Kapitel
Kopp, T., Hendig, I. & Kinkel, S. (2021). Kollaborationsfähige Roboter: Hype oder Zukunftstechnologie? Ergebnisse einer empirischen Marktanalyse. <i>Industrie 4.0 Management</i> , (3), 53–57.	keine Zuordnung (zitiert in 2.2.3.1)
Baumgartner, M., Kopp, T. & Kinkel, S. (2020). Industrielle Mensch-Roboter-Kollaboration in KMU. KMU unterschätzen das Potenzial der Mensch-Roboter-Interaktion. <i>wt Werkstattstechnik online</i> , 110(3), 146–150.	keine Zuordnung (zitiert in 2.2.3.3)

Überdies sind die Inhalte in Kapitel 4.6.2.4 sowie Teile des Anhangs in ähnlicher Form in der betreuten Abschlussarbeit von Baumgartner (2020) enthalten.

In Hinblick auf die Übersetzung von Abbildung 3.6 gilt folgender Haftungsausschluss gemäß den Vorgaben von SAGE Publications:

While every effort has been made to ensure that the contents of this publication are factually correct, neither the authors nor the publisher accepts, and they hereby expressly exclude to the fullest extent permissible under applicable law, any and all liability arising from the contents published in this Article, including, without limitation, from any errors, omissions, inaccuracies in original or following translation, or for any consequences arising therefrom. Nothing in this notice shall exclude liability which may not be excluded by law. Approved product information should be reviewed before prescribing any subject medications.

Obwohl alle Anstrengungen unternommen wurden, um sicherzustellen, dass der Inhalt dieser Publikation sachlich korrekt ist, akzeptieren weder die Autoren noch der Verlag, und sie schließen hiermit ausdrücklich und im größtmöglichen, nach geltendem Recht zulässigen Umfang jede Haftung aus, die sich aus dem in dieser Publikation veröffentlichten Inhalt ergibt, einschließlich, aber nicht beschränkt auf Fehler, Auslassungen, Ungenauigkeiten in der ursprünglichen oder nachfolgenden Übersetzung oder für daraus entstehende Folgen. Nichts in diesem Hinweis schließt eine Haftung aus, die gesetzlich nicht ausgeschlossen werden kann. Genehmigte Produktinformationen sollten vor der Verschreibung der betreffenden Medikamente überprüft werden.

B. Vorstudien

I. Vorstudie 1

a. Ziele von Unternehmen bei Cobot-Einführungen

Steigerung der Arbeitsplatzattraktivität: In den vergangenen Jahren zeigte sich speziell in hochindustrialisierten Ländern wie Japan, den USA oder in vielen europäischen Ländern ein zunehmender Fachkräftemangel (McCarthy, 2019). Dies betrifft auch produzierende Unternehmen, welche Schwierigkeiten haben, ausreichend qualifizierte Ingenieur:innen und Produktionsmitarbeitende zu gewinnen (Giffi et al., 2018). So benötigten Unternehmen im Jahre 2018 durchschnittlich 23 Tage länger um ihre Stellen zu besetzen als in 2015 (Giffi et al., 2018, S. 4). Die Unternehmen konkurrieren demnach auf dem Arbeitsmarkt um geeignetes Personal und investieren mitunter viel Zeit und Geld, um ihre Chancen auf die Gewinnung von Fachkräften zu erhöhen. Das Ansehen und die Bekanntheit großer Unternehmen und deren finanziellen Möglichkeiten, die es ihnen oft erlauben, höhere Gehälter zu bezahlen, erschwert die Situation besonders für KMU. Darüber hinaus leidet insbesondere die verarbeitende Industrie unter einem schlechten Image innerhalb der jüngeren Generation, das durch Berichte über das Offshoring von Fertigungsaktivitäten und die Assoziation solcher Jobs mit typischerweise monotoner, gefährlicher und schmutziger Arbeit als vergleichsweise weniger attraktiv empfunden wird als Büroarbeit (Skevi et al., 2014). Folglich betrachten viele Unternehmen Cobots als ein Mittel, um ihre Mitarbeitenden von körperlich und geistig anstrengenden Aufgaben zu entlasten und damit ihre Arbeitsplatzattraktivität für Bewerber:innen zu erhöhen. Zudem glauben sie, dass Cobots dazu beitragen, ihren Ruf als innovatives Unternehmen zu festigen, was gerade für junge und technologieorientierte Bewerber:innen ein entscheidendes Kriterium darstellen kann (Kopp et al., 2020).

Erhöhung der Flexibilität in der Fertigung: Unternehmen versuchen zunehmend flexible Kundenanforderungen zu erfüllen, ohne die traditionellen Vorteile der Massenproduktion wie eine hohe Kosteneffizienz aufgrund von Skaleneffekten aufzugeben (Wang et al., 2017). Bis dato wurden industrielle Roboter vorwiegend in automatisierten Zellen für die Massenproduktion in großen Losgrößen verwendet (Bruno & Antonelli, 2018). Gerade für KMU, die oft mit kleinen Losgrößen eine große Bandbreite an Produktvarianten produzieren, sind diese vollautomatisierten Roboterzellen nicht ökonomisch, da die Umrüstung zu viel Zeit erfordert (Beckert et al., 2016), die Flexibilität der Produktion einschränkt und es somit erschwert wird, die spezifischen Kundenanforderungen zu erfüllen. Mithilfe einfach umrüstbarer Cobots zielen Unternehmen darauf ab, die Flexibilität ihrer Produktion zu erhöhen, um entsprechend flexibel auf die Kundenanforderungen reagieren zu können.

Erhöhung der Produktivität: Wohingegen die Stärken von Menschen in sensomotorischen Tätigkeiten oder in der Adaption neuer Prozesse liegen, verfügen Roboter über Vorteile in der Exaktheit, Ausdauer und Wiederholgenauigkeit (Grahn, Gopinath, Wang & Johansen, 2018; Krüger, Lien & Verl, 2009; Ranz et al., 2018). In der traditionellen Fertigung mit getrennt arbeitenden Menschen und Robotern ließen sich die jeweiligen Stärken bisher nicht direkt kombinieren. Einige Anwendungsszenarien bergen die Chance, die Durchlaufzeit von Produkten bei gleichbleibendem Personaleinsatz zu reduzieren, z. B. wenn ein Cobot einem Mitarbeitenden die Erfüllung seiner Arbeit erleichtert, indem er Werkstücke in einer ergonomisch günstigen Position hält oder bestimmte Teilarbeitsschritte wie den Abtransport von Werkstücken flexibel übernimmt. Dementsprechend erwarten einige Unternehmen Produktivitätssteigerungen durch den Einsatz von MRI, wenn die Vollautomatisierung eines kompletten Arbeitsschritts nicht umsetzbar oder ineffizient wäre, aber die Übernahme kleiner Arbeitsschritte durch einen Cobot die Arbeit für die Mitarbeitenden erleichtern könnte.

b. Branchenverteilung in Stichprobe

Abbildung A.1 zeigt die Verteilung der Industriebranchen, denen die Unternehmen der Befragten angehören, im Vergleich zur Verteilung in der deutschen Wirtschaft mit Stand 2019. Obwohl einige Branchen wie bspw. der Maschinenbau mit 24.7 % im Vergleich zu 16.9 % in der Stichprobe überrepräsentiert sind, erscheint die Verteilung insgesamt weitestgehend repräsentativ für die deutsche Wirtschaft, was auch durch eine signifikante Spearman-Korrelation gestützt wird, $r = .71$, $p = .020$.

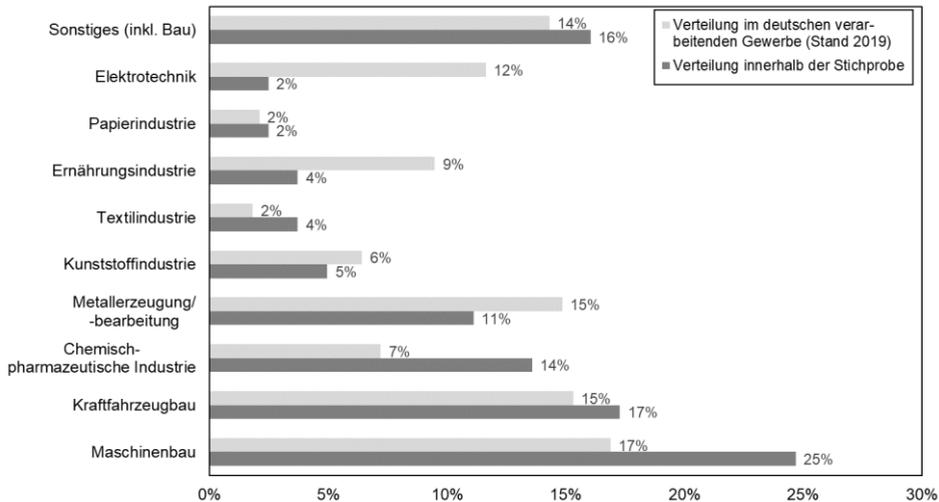


Abbildung A.1: Verteilung der industriellen Sektoren in der Stichprobe und im gesamten deutschen verarbeitenden Gewerbe bezogen auf Beschäftigtenzahlen. Datenquelle: Statistisches Bundesamt Deutschland (2020).

II. Vorstudie 2

a. Interview-Leitfaden

Leitfaden für Experteninterviews AP1 TP 1.2

Interviewer: _____ Datum: _____

Experte: _____ Projektpartner/Firma: _____

Begrüßung/Briefing

- *Ggf. kurze eigene Vorstellung, Verweis auf Projekt ProBot*
- *Ablauf des Interviews: ca. 45-60 Minuten, offene Fragen, flexibler Verlauf, Themenstellung Cobot-Einsatz, Einverständnis zur akustischen Aufnahme*
- *Kurze Einführung zum Cobot-Einsatz (s. Seite 1)*
- *Keine personenbezogene Auswertung, Freigabeprozess bei Verwendung firmenbezogener Daten außerhalb des Projekts*
- *Bedanken für Teilnahme als Experte*

Beginn

0. Einstiegsfrage: „Wir würden zuallererst gerne etwas über Ihren Hintergrund und die gegenwärtige Situation in der Firma [Firmenname] erfahren. Bitte erklären Sie uns kurz in wenigen Sätzen Ihre Rolle/Funktion in Ihrem Unternehmen. Mit was beschäftigen Sie sich hier bei der Firma [Firmenname]? Welche Themen beschäftigen Sie aktuell? Welche Erfahrungen haben Sie bisher mit Robotern bzw. Automation gemacht?“

1. Leitfragen für offenes Gespräch:

1.1. Grundsätzliche Überlegen:

Wenn Sie an die Einführung eines Cobots bzw. dessen Einsatz in Ihrem Unternehmen denken, welche Gedanken und Überlegungen kommen Ihnen dazu in den Sinn?

1.2. Auslöser („Trigger“) für Cobots:

Was könnten Gründe sein, die Ihr Unternehmen veranlassen, über die Einführung eines Cobots nachzudenken?

1.3. Ziele durch Cobots, Pro- und Contra-Punkte:

Welche Ziele verbinden Sie mit dem Einsatz eines Cobots für Ihr Unternehmen? Was könnte für und was gegen einen Cobot-Einsatz sprechen? Inwiefern könnte ein Cobot bisher verfügbaren Automatisierungsansätzen überlegen sein?

- *Produktivitätssteigerung, attraktivere Arbeitsplätze (körperliche Entlastung, geistige Anforderung; für bestehendes Personal, bei Personalwerbung), Fachkräftemangel entgegenwirken, Kostensenkungen, Qualitätssteigerung*

1.4. Gestaltung des Einführungsprozesses:

Angenommen, Ihre Firma überlegt sich nun konkret einen Cobot einzusetzen. Welche Faktoren und Perspektiven sollten im Rahmen des Entscheidungsprozesses einbezogen werden? Wie sollte aus Ihrer Sicht der Einführungsprozess des Cobots gestaltet sein?

- *Was könnten hierbei mögliche Herausforderungen sein?*
- *Für wie kritisch halten Sie die Dauer des Einführungsprozesses? Wie lange sollte dieser im Optimalfall dauern?*
- *Welche Mitarbeiter sollten aus Ihrer Sicht am Einführungsprozess beteiligt sein?*

1.5. Auswirkungen auf Personal:

Welche Auswirkungen hat der Cobot-Einsatz Ihrer Meinung nach in Hinblick auf Ihre Mitarbeiter(-struktur) bzw. Ihr Personalwesen?

- *Für welchen Job/Prozessschritt findet sich aktuell am schwierigsten Personal?*
- *Welche Mitarbeiter werden künftig mit dem Cobot in Kontakt kommen? Wer übernimmt dabei welche Rollen und Aufgaben? (Einrichtung / Anpassung des Cobots)*

1.6. Automatisierungspotenzial einzelner Arbeitsschritte:

Welchen Prozess/Arbeitsschritt würden Sie mithilfe des Cobots am ehesten verändern? Warum gerade dieser Arbeitsschritt? Welche Arbeitsschritte sollten produktiver werden?

- *Wo liegen derzeit Engpässe in der Fertigung? Wo passieren die meisten Arbeitsunfälle bisher? Welche Arbeitsplätze üben Ihrer Meinung nach aktuell die höchste Belastung auf die Mitarbeiter aus?*

2. Optionale Nachfragen zu spezifischen Themen, die im offenen Gespräch noch nicht zur Ansprache kamen:

- 2.1.** Welche Auswirkungen hat der Cobot-Einsatz Ihrer Meinung nach auf neue Kompetenz- und Qualifikationsanforderungen in der Belegschaft?
- 2.2.** Welche arbeitsrechtlichen Herausforderungen sehen Sie durch den CoBot-Einsatz?
- 2.3.** Wie könnte aus Ihrer Sicht die Akzeptanz unter den Mitarbeitern gewährleistet werden?
- 2.4.** Welche Auswirkungen hat der Cobot-Einsatz Ihrer Meinung nach auf betriebswirtschaftliche Kennzahlen?

- 2.5.** Welche Auswirkungen hat der Cobot-Einsatz Ihrer Meinung nach auf die Aufrechterhaltung und Anpassung unternehmensweiter Lern- und Innovationsprozesse wie z.B. KVP in Ihrem Betrieb?

b. Workshop-Konzept

Die jeweils zwei- bis dreieinhalbstündigen Workshops unterteilten sich in die folgenden fünf Phasen:

1. Startphase des Workshops mit Begrüßung und Einführung in das Themenfeld: Das Ziel dieser Phase bestand darin, notwendiges Grundwissen zu vermitteln, um eine qualifizierte Beschäftigung mit der Thematik der Cobot-Einführung zu ermöglichen. Hierzu wurden ein ca. zweiminütiges Video, mündliche Erklärungen sowie ein Miniaturmodell eines *KUKA iiwa* eingesetzt.
2. Erfassung positiver und negativer Assoziationen der Mitarbeitenden gegenüber Cobots: Konkret wurden die Mitarbeitenden dazu aufgefordert, in Kleingruppen zu diskutieren und zu notieren, was sie von der Zusammenarbeit mit einem Cobot erwarten, d. h. welche Vorteile und welche Bedenken sie damit assoziieren. Dabei sollten sich die Teilnehmenden möglichst einen Cobot vorstellen, der komplett nach ihren Vorstellungen gestaltet ist. Die einzelnen Aspekte wurden danach im Plenum vorgetragen und von der Workshop-Leitung auf einer Skala mit den Endpunkten *sehr positiv* und *sehr negativ* grob eingeordnet. Dies hatte zum Ziel, neben der Erarbeitung der diskutierten Aspekte auch deren wahrgenommene Relevanz grob einschätzen zu können, die sich aus der Diskussion im Plenum ergab. Da sich die Teilnehmenden dabei unterschiedlich stark engagierten, lässt sich die Einstufung allerdings nicht als abgestimmter Konsens oder als Mehrheitsentscheid, sondern eher als informelles Stimmungsbild bewerten, zumal den Teilnehmenden eine differenzierte Einordnung mitunter schwerfiel.
3. Analyse der Anforderungen an den Cobot: In dieser Phase wurden die Teilnehmenden aufgefordert, Anforderungen an einen Cobot zu formulieren. Die entsprechende Leitfrage lautete: „Wie muss ein Cobot sein, damit ich gerne mit ihm zusammenarbeite?“ Die Workshop-Leitung verdeutlichte hierbei, dass sich die Anforderungen auf alle möglichen Aspekte des Cobots samt dessen Gestaltung und Funktionsweise beziehen können. Anschließend wurden die Aspekte erneut im Plenum gesammelt und von der Workshop-Leitung strukturiert auf einer Skala von *sehr gerne* bis *nicht gerne* eingeordnet.
4. Analyse der Anforderungen an den Einführungsprozess eines Cobots: In dieser Phase wurden die Teilnehmenden aufgefordert, sich in das Szenario zu versetzen, dass ihr Unternehmen die Einführung des Cobots bereits beschlossen hat und sich nun mit der Ausgestaltung des Einführungsprozesses auseinandersetzt. Die

entsprechende Leitfrage lautete: „Wie sollte der Einführungsprozess Ihrer Meinung nach gestaltet sein?“ Die Ergebnisse wurden im Plenum diskutiert und von der Workshop-Leitung direkt in übersichtlicher Form digital notiert. Die Notizen wurden projiziert, sodass am Ende ein von allen Anwesenden mitgetragener Konsens entstand.

5. Endphase des Workshops mit Verabschiedung der Mitarbeitenden: Zum Schluss erhielten die Teilnehmenden die Gelegenheit Feedback abzugeben. Die Workshops wurden von der überwiegenden Mehrheit als positiv eingestuft und die Partizipationsmöglichkeit geschätzt. Anschließend dankte die Workshop-Leitung den Teilnehmenden und verabschiedete sich.

C. Experimentalstudien

I. Framing-Text: Menschenähnlichkeit

a. Maschinelle Beschreibung

Stellen Sie sich vor, Sie werden künftig in Ihrem Unternehmen mithilfe des kollaborierenden Roboters UR-5 arbeiten.

Der UR-5 wurde technisch so ausgestattet, dass Menschen direkt mit dem Roboter zusammenarbeiten können. Mithilfe entsprechender Sensoren des UR-5 wird während der gesamten Zusammenarbeit überwacht, dass dieser nicht mit Ihnen zusammenstößt und Sie sich nicht verletzen können. Durch die einprogrammierte geringe Maximalgeschwindigkeit ist sichergestellt, dass unvorhergesehen Zusammenstöße für Sie nicht schmerzhaft sein können. Die Sensoren des Roboters würden dies auch sofort detektieren und den UR-5 stoppen. Dadurch ist es möglich, den Roboter in unmittelbarer Nähe zum Menschen zu betreiben, ohne dass dieser durch einen Schutzzaun abgetrennt werden muss.

Der UR-5 verfügt nicht über eine Sprachausgabe. Wenn etwas nicht funktioniert, erzeugt der Roboter sofort eine Meldung, damit ein Mitarbeiter die Störung beheben kann. Danach arbeitet der UR-5 ohne Einschränkungen weiter.

Bevor der Roboter in Ihrem Unternehmen aufgebaut wurde, wurde dieser zuvor schon in anderen Unternehmen verwendet. Zuletzt befand er sich bei der Meyer GmbH in Karlshofen im Betrieb. Bei der Meyer GmbH wurde der Roboter dazu eingesetzt, Lasten von bis zu 5kg Gewicht anzuheben. Mit einer Reichweite von bis zu 0,85 Metern konnte der Roboter außerdem dazu verwendet werden, weiter entfernte Bauteile zu greifen.

Aktuell hat der UR-5 noch eine geringe Fehlerrate von ca. 5%. Allerdings ist die programmierte Logik selbstlernend und entwickelt sich weiter, sodass mit der Zeit immer weniger Fehler entstehen werden. Wenn die Mitarbeiter eine Frühstückspause einlegen, verbleibt der UR-5 so lange im Stillstandsmodus, bis die Mitarbeiter den Roboter wieder einschalten. Der UR-5 muss nur einen Tag pro Jahr gewartet werden, um weitere 365 Tage durchgehend betrieben werden zu können.

b. Menschenähnliche Beschreibung

Stellen Sie sich vor, Sie werden künftig in Ihrem Unternehmen gemeinsam mit dem kollaborierenden Roboter Paul zusammenarbeiten.

Paul hat gelernt, sich so zu verhalten, dass er direkt mit Menschen zusammenarbeiten kann. Mithilfe seines Sehvermögens achtet Paul während der gesamten Zusammenarbeit darauf,

dass er nicht mit Ihnen zusammenstößt und Sie sich nicht verletzen können. Durch seine gemächlichen Bewegungen ist sichergestellt, dass unvorhergesehene Zusammenstöße für Sie nicht schmerzhaft sein können. Paul würde diese auch sofort bemerken und bliebe dann stehen. Dadurch ist es möglich, dass Paul in unmittelbarer Nähe zum Menschen arbeitet, ohne dass er hinter einem Schutzzaun weggesperrt werden muss.

Sprechen kann Paul nicht. Wenn etwas nicht funktioniert, meldet er sich sofort, damit ein Mitarbeiter ihm helfen kann. Danach arbeitet Paul ohne Verdruss weiter.

Bevor Paul in Ihr Unternehmen gekommen ist, hat er zuvor schon andere Unternehmen von innen gesehen. Zuletzt verrichtete er seinen Dienst bei der Meyer GmbH in Karlshofen. Bei der Meyer GmbH unterstützte Paul die Mitarbeiter dabei, Lasten von bis zu 5kg Gewicht anzuheben. Mit seiner Reichweite von bis zu 0,85 Metern kann Paul außerdem dabei helfen, weiter entfernte Bauteile zu greifen.

Wie auch seine menschlichen Kollegen, macht Paul aktuell in ca. 5% der Fälle noch Fehler. Allerdings ist Paul lernfähig und entwickelt sich weiter, sodass er mit der Zeit immer weniger Fehler machen wird. Wenn die Mitarbeiter eine Frühstückspause einlegen, kommt Paul zwar nicht mit, wartet aber artig so lange im Stillstand, bis die Mitarbeiter ihn wiederbeleben. Paul braucht nur einen Ruhetag pro Jahr, an dem er fit gemacht wird, um weitere 365 Tage durchzuarbeiten.

II. Framing-Text: Mensch-Cobot-Relation

a. Konkurrenzorientiert

Anmerkung: Der Robotername wurde jeweils gemäß der vorhergehenden Bedingung angepasst.

Der Geschäftsführer der Meyer GmbH berichtet folgendermaßen über seine Erfahrungen mit [Robotername]:

„Die sinnvolle Nutzung der Stärken des Roboters bringt deutliche Vorteile. Durch seinen Einsatz konnten wir unseren Ausschuss deutlich reduzieren. Manche Mitarbeiter fühlten sich durch [Robotername] unter Druck gesetzt, sodass sie ihre Arbeitsgeschwindigkeit so erhöhten, dass sie mit [Robotername] mithalten konnten. Der Roboter ist - im Zusammenspiel mit dem Menschen - für ein zukunftsorientiertes Unternehmen unverzichtbar. In Zukunft werden wir bei manuellen Tätigkeiten verstärkt Roboter wie [Robotername] einsetzen.“

Bei Ihrer zukünftigen Arbeit wird [Robotername] diejenigen Teile Ihrer Aufgaben übernehmen, bei denen Sie dem Roboter unterlegen sind. Durch die bessere Wiederholgenauigkeit und die höhere Geschwindigkeit von [Robotername] verspricht sich Ihr Unternehmen, eine bessere Performance im Vergleich zu einer rein manuellen Montage zu erzielen. Sie werden weiterhin diejenigen Aufgaben übernehmen, die sich mit [Robotername] nicht sinnvoll umsetzen lassen.

b. Kooperationsorientiert

Anmerkung: Der Robotername wurde jeweils gemäß der vorhergehenden Bedingung angepasst.

Der Geschäftsführer der Meyer GmbH berichtet folgendermaßen über seine Erfahrungen mit [Robotername]:

„Die sinnvolle Kombination der Stärken von Mensch und Roboter bringt deutliche Vorteile. Durch den Einsatz des Mensch-Roboter-Teams konnten wir unseren Ausschuss deutlich reduzieren. Manche Mitarbeiter fühlten sich durch das neue Teammitglied motiviert, sodass sie ihre Arbeitsgeschwindigkeit so erhöhten, dass sie mit [Robotername] auf Augenhöhe arbeiten konnten. Der Mensch ist - im Zusammenspiel mit dem Roboter - für ein zukunftsorientiertes Unternehmen unverzichtbar. In Zukunft werden wir den Menschen bei manuellen Tätigkeiten verstärkt Roboter wie [Robotername] zur Seite stellen.“

Bei Ihrer zukünftigen Arbeit wird Sie [Robotername] bei denjenigen Teilen Ihrer Tätigkeiten unterstützen, bei welchen der Roboter besonders nützlich ist. Durch die bessere Wiederholgenauigkeit und die höhere Geschwindigkeit von [Robotername] verspricht sich Ihr Unternehmen, dass Sie dadurch in die Lage versetzt werden, eine bessere Performance im Vergleich zu einer rein manuellen Montage erzielen. Sie werden weiterhin diejenigen Aufgaben übernehmen, bei welchen Ihnen [Robotername] nicht sinnvoll helfen kann.

III. Kodierleitfaden für wahrgenommene Menschenähnlichkeit

		Wahrgenommene Menschenähnlichkeit				menschenähnlich	
		1	2	3	4	5	
Bewertungskriterien	maschinell	1	2	3	4	5	
		Technische Eigenschaften des Cobots werden explizit hervorgehoben. Der Cobot wird explizit als „typisch“ für eine Maschine / einen Roboter beschrieben.	Implizite Nutzung von (eher) maschinenähnlichen Begriffen. Der Cobot wird in seinen Eigenschaften vom Menschen abgegrenzt.	Keine erkennbaren Tendenzen in eine der besagten Richtungen. Widersprüchliche Aussagen.	Implizite Nutzung von (eher) menschenähnlichen Begriffen/ Eigenschaften. Technische Eigenschaften des Cobots werden vermenschlicht	Cobot wird explizit als vergleichbar zu einem Menschen dargestellt. Es werden Parallelen (keine Unterscheidungen!) zwischen einem Menschen und dem Cobot gezogen.	
Beispiele		"mechanisch", "Technik" ...	"festmontiert", "automatisiert", "programmiert", "Greifer", "Wartung", ...	"Roboter" (weder maschinenähnlich noch menschenähnlich)	"reden", "hilfsbereit", "überforder", "denk", "selbstständiges Lernen", Endung -er "Helfer", "Begleiter"	"wie ein Mensch", "menschenähnlich", "wie ein Arbeitskollege"...	
		"Maschine", "Werkzeug"/ "Hilfsmittel", "typischer Industrieroboter", "künstlich"	"im Gegensatz zum Menschen"	"Gewissenhaft (menschenähnlich), aber hat im Gegensatz zum Menschen kein Gehirn (maschinenähnlich)"	"Hand" (statt Greifer), "Pause" (statt Wartung)	"handelt wie auch ein Mensch"	
Zusatzregel	Werden zur Beschreibung des Cobot menschenähnliche Eigenschaften in registrierter Form benutzt, so wird der Mittelwert (3) gewählt. (Beispiel: "Der Cobot kann nicht sprechen.")						

Abbildung A.2: Kodierleitfaden für wahrgenommene Menschenähnlichkeit.

IV. Kodierleitfaden für wahrgenommene Kooperationsorientierung

	Wahrgenommene Kooperationsorientierung				kooperierend
	1 konkurrierend	2	3	4	
Bewertungskriterien	Es werden Bedenken bezüglich des Verlusts des Arbeitsplatzes geäußert. Es werden Bedenken bezüglich Verletzungen und Sicherheitsrisiken geäußert.	Es werden mögliche Bedenken hinsichtlich des Nutzens (Schwächen/ Einschränkungen des Cobots) geäußert. Es werden mögliche Bedenken hinsichtlich schlechterer Arbeitsumstände (Stress etc.) geäußert.	Keine erkennbaren Tendenzen in eine der besagten Richtungen.	Die unterstützende Funktion des Cobots wird erwähnt.	Die unterstützende Funktion des Cobots wird befürwortet (positive Wertung).
Beispiele	"könnte Arbeitsplatz übernehmen", "Menschen könnten in Zukunft ersetzt", ... "könnte zu Verletzungen führen"	"Für Aufgabe xy braucht es weiterhin einen Menschen", "Roboter sind nicht gut bei Aufgaben wie..." "Mensch muss dadurch schneller arbeiten"	"Roboter werden eingesetzt um schneller zu werden"	Es werden konkrete Aufgaben genannt, bei denen der Cobot unterstützen kann. "hilft", "hilfsbereit", "ergänzt", "Kollege", "erleichtert", ...	Die Versuchsperson erkennt den persönlichen Nutzen des Cobots für bestimmte Aufgaben. "Unterstützung ist sinnvoll", "Unterstützung gefällt mir", ... "kann mir lästige/ körperliche Aufgaben abnehmen"
Zusatzregel	Bei Aussagen der Versuchspersonen, die zwei gegensätzliche Argumente gegenüberstellen (z. B. hilfsbereit, aber macht zu viele Fehler), wird die Mittelkategorie (3) gewertet. Ausnahme: Äußert die Versuchsperson Ängste vor Arbeitsplatzverlusten oder Verletzungen, wird immer mit '1' bewertet (dominierender Faktor).				

Abbildung A.3: Kodierleitfaden für wahrgenommene Kooperationsorientierung.

V. Box-Plots zur Manipulationskontrolle in Experimentalstudie 1

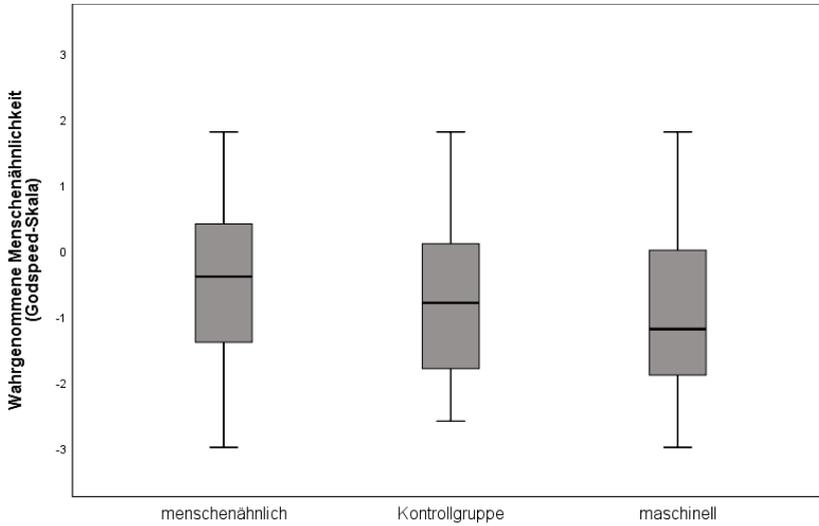


Abbildung A.4: Verteilung der Daten zur wahrgenommenen Menschenähnlichkeit (gemessen anhand der *Godspeed*-Skala) je Versuchsgruppe in Experimentalstudie 1 als Box-Plot.

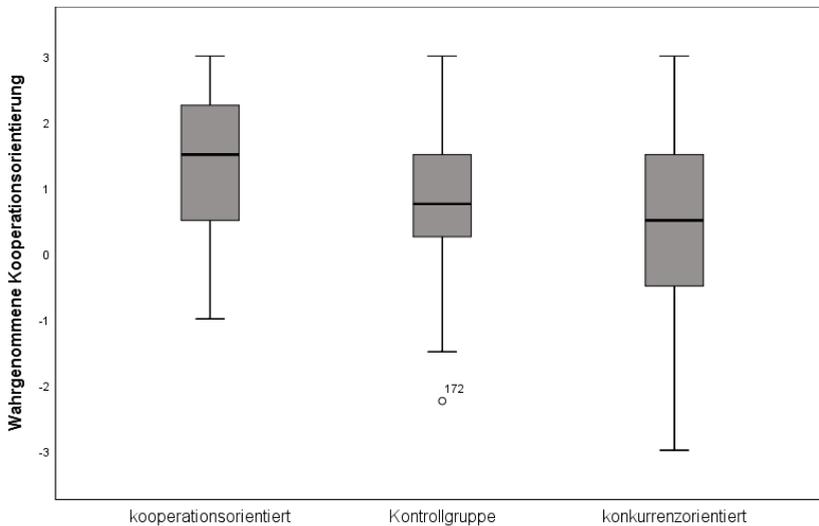


Abbildung A.5: Verteilung der Daten zur wahrgenommenen Kooperationsorientierung je Versuchsgruppe (gemessen anhand eigener Ad-hoc-Skala) in Experimentalstudie 1 als Box-Plot.

VI. Box-Plot zur Manipulationskontrolle in Experimentalstudie 2

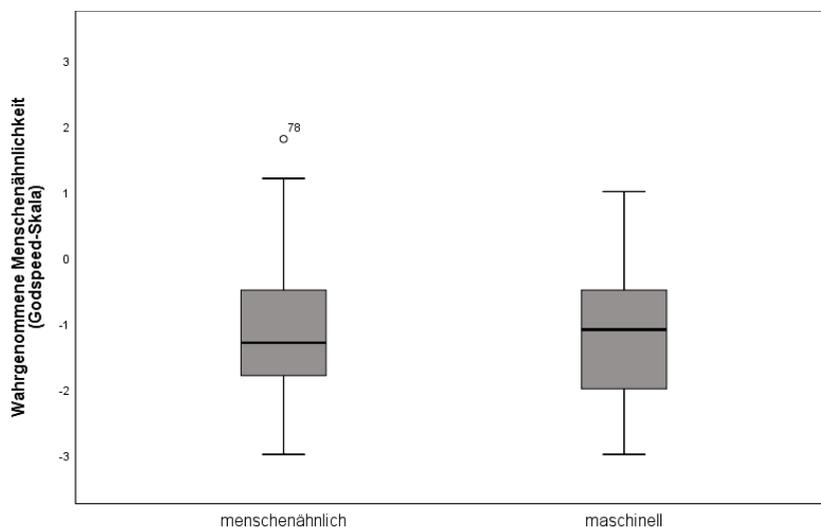


Abbildung A.6: Verteilung der Daten zur wahrgenommenen Menschenähnlichkeit (gemessen anhand der *Godspeed*-Skala) je Versuchsgruppe in Experimentalstudie 2 als Box-Plot.

VII. Übersicht Akzeptanz-Items

Tabelle 43: Übersicht über die in den Experimentalstudien verwendeten Akzeptanz-Items.

Konstrukt	Item	Herkunft	Exp. 1	Exp. 2
Selbstwirksamkeit	Nach einer entsprechenden Einweisung wäre ich in der Lage, einen Roboter wie [Robotername] zu benutzen.	TAM 3		✓
Arbeit am gemeinsamen Arbeitsplatz	Es würde mir nichts ausmachen, wenn [Robotername] mit mir an einem gemeinsamen Arbeitsplatz arbeitet.	Bröhl et al. (2017) als Konstrukt <i>rechtliche Implikation</i>		✓
Bereitschaft zum Berichten von Ergebnissen	Ich hätte keine Schwierigkeiten damit, anderen von den	TAM 2		✓

Konstrukt	Item	Herkunft	Exp. 1	Exp. 2
	Ergebnissen des [Robotername] zu berichten.			
Ergebnisqualität	Die Ergebnisqualität des [Robotername] wäre hoch.	TAM 2		✓
Wahrgenommenes Vergnügen	Ich hätte bei der Nutzung des [Robotername] Spaß.	TAM 3		✓
Angst vor sozialer Isolation	Ich würde befürchten, dass ich durch [Robotername] den Kontakt zu meinen Kollegen verliere.	Bröhl et al. (2017) als Konstrukt <i>soziale Implikation</i>		✓
(Generelle) Roboterangst	In der Gegenwart von [Robotername] würde ich mich unwohl fühlen.	TAM 3		✓
Wahrgenommene Benutzerfreundlichkeit	Die Interaktion mit [Robotername] wäre einfach.	TAM	✓	✓
Wahrgenommener Nutzen	[Robotername] wäre für meine Arbeit nützlich.	TAM	✓	✓
Nutzungsintention	Wenn ich wählen könnte, ob mich bei meiner Arbeit unterstützen soll, würde ich mich für die Unterstützung durch [Robotername] entscheiden.	TAM	✓	✓
Wahrgenommene Sicherheit	Bei der Nutzung von [Robotername] würde ich mich sicher fühlen.	Bröhl et al. (2017)	✓	✓
Tatsächlicher Gebrauch	Ich würde [Robotername] gegenüber anderen Maschinen im industriellen Umfeld bevorzugen.	TAM	✓	✓
Angst vor Jobverlust	Ich würde befürchten, dass ich durch [Robotername] langfristig durch meinen Job verlieren werde.	Bröhl et al. (2017) als Konstrukt <i>ethische Implikation</i>	✓	✓
Einstellung zur Nutzung I	Die Verwendung des [Robotername] ist eine gute Idee.	UTAUT	✓	
Einstellung zur Nutzung II	Mir würde es gefallen mit dem [Robotername] zu arbeiten	UTAUT	✓	

Konstrukt	Item	Herkunft	Exp. 1	Exp. 2
Willkommener Arbeitskollege	Ich hätte den [Robotername] gerne als Arbeitskollegen.	angelehnt an Weiss et al. (2008)	✓	

VIII. Reliabilität verwendeter Skalen

a. Reliabilitäten in Experimentalstudie 1

Die interne Konsistenz sämtlicher Skalen wurde anhand Cronbachs α gemessen. Dabei ist zu berücksichtigen, dass Cronbachs α in Abhängigkeit der Anzahl an Items der Skala unterschiedlich zu interpretieren ist, da eine höhere Item-Anzahl mathematisch zu höheren Werten führt. Gleichsam ist darauf hinzuweisen, dass die Voraussetzungen für die Bestimmung von Cronbachs α in der Praxis häufig nicht gegeben sind, wenn – wie im vorliegenden Fall – nur eine einmalige Testung vorgenommen wird. Dies kann u. U. zu einer Überschätzung der internen Konsistenz führen (Döring & Bortz, 2016, S. 468f.).

Mit Ausnahme der *Godspeed*-Skala erreichten alle Skalen gemäß der Faustregel einen guten Wert ($\alpha > .80$). Die Reliabilität der *Godspeed*-Skala lässt sich als akzeptabel ($\alpha > .70$) bezeichnen. Die Items zur Akzeptanzmessung weisen eine exzellente interne Konsistenz ($\alpha > .90$) auf.

Tabelle 44: Reliabilitätswerte der verwendeten Skalen in Experimentalstudie 1.

Skala	Anzahl Items	Cronbachs α	Quelle
<i>Godspeed</i> -Skala für wahrgenommene Menschenähnlichkeit	5	.770	Ho und MacDorman (2017)
Skala für wahrgenommene Kooperationsorientierung	4	.854	Selbst erstellt
NARS	14	.827	Nomura et al. (2004)
TAS (Vertrauen, Pre-Framing)	5	.812	Pöhler et al. (2016)
TAS (Misstrauen, Pre-Framing)	6	.864	Pöhler et al. (2016)
TAS (Vertrauen, Post-Framing)	5	.828	Pöhler et al. (2016)
TAS (Misstrauen, Post-Framing)	6	.849	Pöhler et al. (2016)
Akzeptanz	9	.905	Items aus verschiedenen Quellen

b. Reliabilitäten in Experimentalstudie 2

Analog zu Experimentalstudie 1 wurde die interne Konsistenz sämtlicher Skalen anhand Cronbachs α gemessen. Dabei ergaben sich wesentlich geringere Werte, welche die Reliabilität einiger Skalen in Frage stellen, allerdings auch vor dem Hintergrund der deutlich geringeren Stichprobe mit Vorsicht zu interpretieren sind. Besonders niedrig fiel Cronbachs α mit .335 bei der Mensch-Roboter-Relation aus. Bei der Mensch-Roboter-Relation zeigte sich nur zwischen den ersten beiden Items eine hohe Korrelation von $r = .496$; alle weiteren Korrelationen liegen unter $r < .200$. Diese Skalen werden bei den folgenden Auswertungen nicht zu einem Wert zusammengefasst, sondern je Item ausgewertet. Auch andere Skalen wie z. B. die NARS mit einer relativ hohen Anzahl an Items erreichte nur einen fragwürdigen ($\alpha > .60$) Reliabilitätswert. Bei der Itemanalyse fällt insbesondere das Item „Das Wort Roboter hat keine Bedeutung für mich“ mit einem negativen Wert für die Trennschärfe auf. Dies lässt sich dadurch begründen, dass die Studierenden in ihren Vorlesungen alle schon theoretisch mit Robotern in Kontakt gekommen sind und dieses Item daher hoch bewerteten. Die NARS13-Skala ohne dieses Item erreicht einen Wert für Cronbachs $\alpha = .661$.

Tabelle 45: Reliabilitätswerte der verwendeten Skalen in Experimentalstudie 2.

Skala	Anzahl Items	Cronbachs α	Quelle
<i>Godspeed</i> für wahrgenommene Menschenähnlichkeit	5	.644	Ho und MacDorman (2017)
Skala für wahrgenommene Kooperationsorientierung	4	.335	Eigene Skala
NARS	14	.602	Nomura et al. (2004)
TAS (Vertrauen, Pre-Framing)	5	.631	Pöhler et al. (2016)
TAS (Misstrauen, Pre-Framing)	6	.850	Pöhler et al. (2016)
TAS (Vertrauen, Post-Framing)	5	.646	Pöhler et al. (2016)
TAS (Misstrauen, Post-Framing)	6	.696	Pöhler et al. (2016)
TAS (Vertrauen, Post-Interaktion)	5	.650	Pöhler et al. (2016)
TAS (Misstrauen, Post-Interaktion)	6	.793	Pöhler et al. (2016)
Akzeptanz	13	.660	Items aus verschiedenen Quellen
Akzeptanz kurz (Wahrgenommene Nützlichkeit, wahrgenommene Benutzerfreundlichkeit, Nutzungsabsicht)	3	.621	Items aus verschiedenen Quellen
Kontrollüberzeugung im Umgang mit Technik (KUT)	4	.761	Neyer et al. (2012, S. 88)

IX. Experimentalstudie 2: Manipulationskontrolle nach Art des Studiengangs

Tabelle 46: Ergebnisse der Manipulationskontrolle für wahrgenommene Menschenähnlichkeit je nach Studiengang (technisch / nicht-technisch) der Versuchsteilnehmenden.

	<i>n</i>	Selbstauskunft anhand <i>Godspeed</i> -Skala			Freitext-Bewertung		
		<i>M</i>	<i>SD</i>	<i>t</i> -Test	<i>M</i>	<i>SD</i>	<i>t</i> -Test
<i>Technischer Studiengang</i>							
Maschinell	19	-1.34	1.01	$t(35) = -1.481$ $p = .148$	2.79	0.86	$t(35) = -0.150$ $p = .882$
Menschen-ähnlich	18	-0.81	1.19		2.83	0.92	
<i>Nicht-technischer Studiengang</i>							
Maschinell	17	-0.96	1.00	$t(29) = 1.436$ $p = .162$	3.00	0.61	$t(29) = -0.538$ $p = .595$
Menschen-ähnlich	14	-1.14	0.68		3.14	0.86	

X. Experimentalstudie 2: Konstruktionsanleitung

Die Konstruktionsanleitung findet sich auch in der betreuten Abschlussarbeit von Baumgartner (2020, S. XVII).

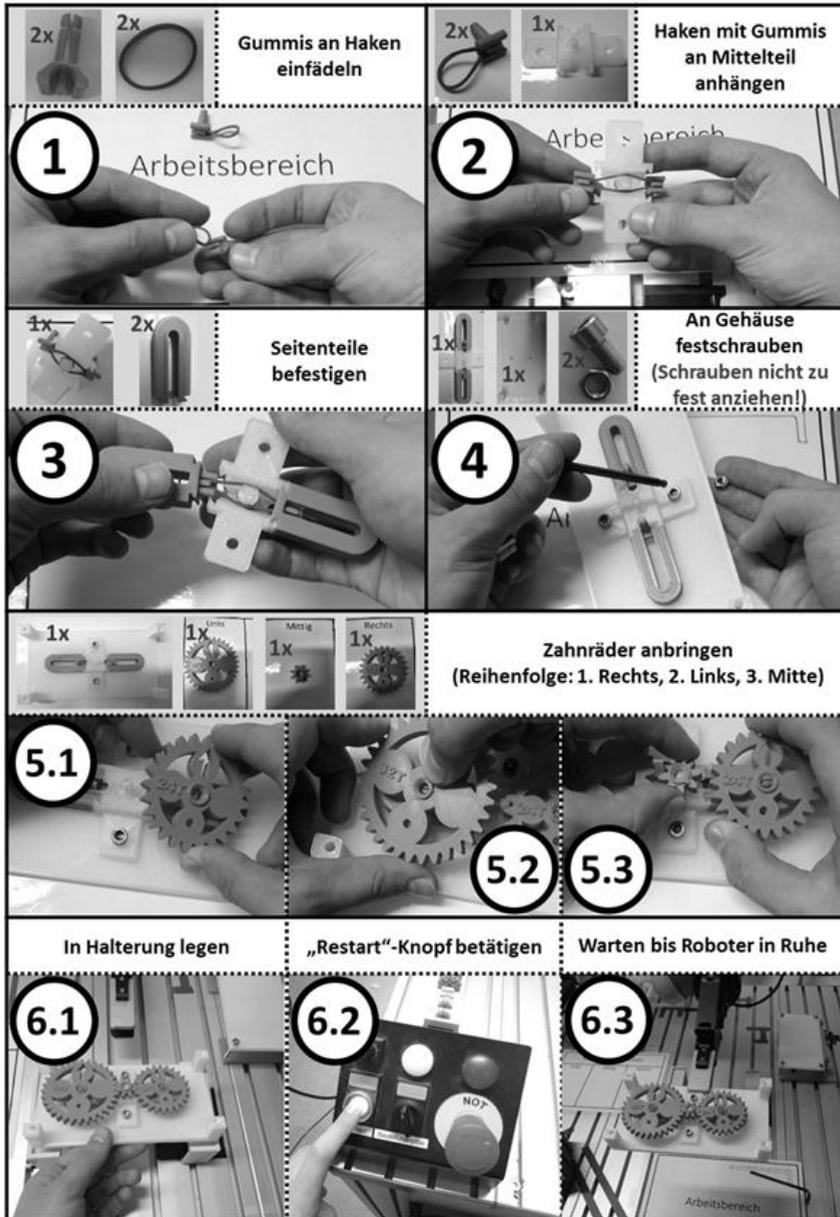




Abbildung A.7: Konstruktionsanleitung zur Montage der Zahnradgetriebe in Experimentalstudie 2.



Im Fokus dieser Arbeit steht das Vertrauen (und Misstrauen) von Mitarbeitenden in kollaborationsfähige Roboter (sog. Cobots) am industriellen Arbeitsplatz. Der empirische, interdisziplinäre und anwendungsnahe Forschungsansatz greift auf Theorien aus verschiedenen Disziplinen und auf quantitative sowie qualitative Untersuchungsmethoden zurück. Die Stichproben umfassen Mitarbeitende auf operativer und leitender Ebene in produzierenden Unternehmen sowie Studierende. Die Ergebnisse verdeutlichen u. a. den signifikanten Einfluss des sprachlichen Framings auf das Vertrauen der Produktionsmitarbeitenden. Das Framing bezieht sich dabei einerseits auf die Vermenschlichung eines Cobots und andererseits auf die wahrgenommene Mensch-Cobot-Relation im Spannungsfeld zwischen Kooperation und Konkurrenz. Ein vertrauensförderlicher Effekt der Vermenschlichung stellt sich ein, wenn sich die Mitarbeitenden in einer kooperierenden Relation zum Cobot sehen. Ferner beeinflussen näher zu untersuchende personenspezifische und kontextuelle Faktoren die Wirkkraft des Framings. Vertrauen und Misstrauen erscheinen letztlich als konzeptionell unterschiedliche, multidimensionale und sich zeitdynamisch entwickelnde Konstrukte. Daraus ergeben sich unternehmerische und gesellschaftliche Implikationen auch in Hinblick auf ähnliche Technologien und Anwendungskontexte sowie Bedarfe für anknüpfende anwendungsnahe und theoriebildende Forschungsarbeiten.

