Head and Heart:
On the Acceptability of Sophisticated Robots Based on an Enhancement of the Mind Perception Dichotomy and the Uncanny Valley of Mind

Inaugural-Dissertation

zur Erlangung der Doktorwürde (Dr. rer. nat.) der

Fakultät für Humanwissenschaften

der

Julius-Maximilians-Universität Würzburg

vorgelegt von

Andrea Grundke, Master of Science

aus Lohr am Main

Würzburg, 2023

Erstgutachter:	Prof. Dr. Markus Appel, Julius-Maximilians-Universität Würzburg
Zweitgutachterin:	Prof. Dr. Birgit Lugrin, Julius-Maximilians-Universität Würzburg
Drittgutachter:	Prof. Dr. Jan-Philipp Stein, Technische Universität Chemnitz

Tag der Disputation:	18. Oktober 2023

Acknowledgments

Summary (German)

Mit der stetigen Weiterentwicklung von künstlicher Intelligenz besteht das Bestreben, das ausgedrückte Bewusstsein von Robotern dem menschlichen Bewusstsein immer mehr nachzubilden. Ebenso wie ein sehr menschenähnliches Äußeres zu Aversion gegenüber solchen Robotern führen kann, hat neuere Forschung dargelegt, dass auch das augenscheinlich von Maschinen ausgedrückte Bewusstsein für negative Bewertungen verantwortlich sein kann. Ziel dieser Arbeit ist es, Facetten der durch Maschinen mit menschenähnlichem Bewusstsein evozierten Aversion (*Uncanny Valley of Mind*) mit drei empirischen Forschungsprojekten aus psychologischer Perspektive in verschiedenen Kontexten inklusive der daraus resultierenden Konsequenzen zu erforschen. Damit sollen der bisherige Kenntnisstand auf diesem Gebiet erweitert und Implikationen aufgezeigt werden.

In dem ersten Projekt wird die Perspektive bisheriger Arbeiten in dem Forschungsgebiet umgekehrt und mit zwei Online-Experimenten (N_1 = 335, N_2 = 536) gezeigt, dass Menschen Unheimlichkeit gegenüber Robotern empfinden, die scheinbar menschliche Gedanken lesen können und somit sogar mentale Fähigkeiten aufzuweisen scheinen, die Menschen selbst in sozialen Interaktionen nicht zur Verfügung stehen. Im Vergleich dazu ist die Aversion gegenüber Robotern, die menschliche Emotionen lesen, niedriger ausgeprägt. Dieses Fähigkeit ist bereits aus der Mensch-Mensch-Interaktion bekannt und das Ergebnismuster erweist sich als stabil ungeachtet der Inbezugnahme verschiedener Persönlichkeitsmerkmale (HEXACO). Das zweite Forschungsprojekt nutzt in zwei Online-Experimenten (N_1 = 559, N_2 = 396) verschiedene Video-Stimuli um herauszufinden, ob Empathie für einen von einem Menschen verletzt werdenden Roboter eine Möglichkeit ist, das Uncanny Valley of Mind abzuschwächen. Ein hervorzuhebendes Resultat dieser Arbeit ist, dass Aversion in dieser Studie nicht wie präregistriert durch die Manipulation der Bewusstseinszustände des Roboters aufgekommen ist, sondern durch dessen attribuierte Inkompetenz und sein Versagen. Bei der Bewertung des Roboters scheint es eine untergeordnete Rolle zu spielen, dass die Inkompetenz des Roboters durch die schädigenden Handlungen des Menschen hervorgerufen wurde. Um die Ergebnisse der ersten beiden Projekte in Beziehung zu setzen und die methodische Vielfalt der Arbeit durch Live-Interaktionen zu erweitern, wurde das dritte Projekt als Labor-Experiment (N_1 = 104, Interaktion mit dem Roboter NAO) und aufbauend als Online-Experiment (N_2 = 589, Interaktion mit einer simulierten künstlichen Intelligenz) konzipiert. Es beschäftigt sich mit der Frage, ob Menschen sich in ihrem Status bedroht fühlen, wenn sie mit einer Maschine zusammenarbeiten, die arbeitsrelevante Aufgaben besser als der Mensch erledigen kann. Diese Hypothese wird durch die Daten bestätigt. Entgegen den Hypothesen zeigen die Ergebnisse des in den Arbeitskontext eingebetteten Projekts ferner, dass die Statusbedrohung einen positiver Prädiktor für die Bereitschaft zur Interaktion darstellt. Die Nützlichkeit der Maschine ist allerdings der stärkste Prädiktor der Bereitschaft zur Interaktion. Die Nützlichkeit erklärt zudem das Aufkommen der Statusbedrohung.

Mit Blick auf die Ergebnisse der drei Projekte lässt sich resümieren, dass Menschen recht positiv auf Maschinen mit menschenähnlichen mentalen Fähigkeiten reagieren, sobald erklärende Variablen und konkrete Szenarien mitberücksichtigt werden. Solange Menschen die Nützlichkeit der Maschinen deutlich gemacht wird, diese aber nicht vollkommen autonom sind, scheinen Menschen gewillt zu sein, mit diesen zu interagieren und dabei Gefühle von Aversion zugunsten der antizipierten Vorteile in Kauf zu nehmen.

Summary (English)

With the continuous development of artificial intelligence, there is an effort to let the expressed mind of robots resemble more and more human-like minds. However, just as the human-like appearance of robots can lead to feelings of aversion to such robots, recent research has shown that the apparent mind expressed by machines can also be responsible for their negative evaluations. This work strives to explore facets of aversion evoked by machines with human-like mind (*uncanny valley of mind*) within three empirical projects from a psychological point of view in different contexts, including the resulting consequences. The goal is to expand the current knowledge in this area and to highlight possible implications.

The first research project reverses the perspective of previous work in the research area. It consists of two online experiments (N_1 = 335, N_2 = 536) showing that humans feel eeriness in response to robots that can read human minds and thus have even more mental skills than humans can muster in social interactions. In comparison, the aversion to robots that read human emotions, a capability people are familiar with from human-human interaction, is lower. Moreover, these findings are stable regardless of various personality variables (HEXACO). The second research project uses several video stimuli in two online experiments (N_1 = 559, N_2 = 396) to determine whether empathy for a robot being harmed by a human is a way to alleviate the uncanny valley of mind. A result of this work worth highlighting is that aversion in this study did not arise from the manipulation of the robot's mental capabilities as pre-registered but from its attributed incompetence and failure. In evaluating the robot, it seems to play a minor role that the human's harmful actions caused the robot's vulnerability and thereby implied incompetence. In order to connect the results from the first two projects and to enrich the methodological variety of experiments with live interactions, the third project is designed as a laboratory experiment (N_1 = 104, interaction with the robot NAO) and afterward replicated as an online experiment (*N_2* = 589, interaction with a simulated artificial intelligence). This third project examines whether people feel status-threatened if they work with a machine that can perform job-relevant tasks better than humans. This hypothesis is confirmed by the data. Contrary to the expectations, the results of the project embedded in the work context show that status threat predicts the willingness to interact positively. The results further demonstrate that the machine's perceived usefulness is the strongest predictor of the willingness to interact with it. Moreover, the perceived usefulness explains the emergence of status threat.

The results of the three projects highlight that people will react fairly positively to machines with human-like mental capabilities if explanatory variables and concrete scenarios are considered. As long as the machine’s usefulness is palpable to people, but machines are not fully autonomous, people seem willing to interact with them, accepting aversion in favor of the expected benefits.

LIST OF ABBREVIATIONS

AI	Artificial Intelligence
ANOVA	Analysis of Variance
APA	American Psychological Association
B.C.E.	Before Common Era
C.E.	Common Era
DOI	Digital Object Identifier
GPT	Generative Pre-Trained Transformers
HEXACO	Honesty-Humility, Emotionality, Extraversion, Agreeableness, Conscientiousness, Openness to Experience
I-S-T 2000R	Intelligenz-Struktur-Test 2000R
MANOVA	Multivariate Analysis of Variance
MTURK	Amazon Mechanical Turk
NARS	Negative Attitude Toward Robots Scale
OSF	Open Science Framework
SPSS	Statistical Package for the Social Sciences
USD	US-Dollar
WEIRD	White, Educated, Industrialized, Rich, Democratic

List of Included Projects

The publications forming this cumulative dissertation have been published in peer-reviewed scientific journals (Projects 1 and 2) or are an advance online publication in a peer-reviewed scientific journal (Project 3). The following list provides the formal references for all included articles, including their digital object identifiers (doi).

Project 1: Grundke, A., Stein, J.-P., & Appel, M. (2022a). Mind-reading machines: Distinct user responses to thought-detecting and emotion-detecting robots. *Technology, Mind, and Behavior, 3*(1), 1-12. https://doi.org/10.1037/tmb0000053

Project 2: Grundke, A., Stein, J.-P., & Appel, M. (2023). Improving evaluations of advanced robots by depicting them in harmful situations. *Computers in Human Behavior, 140,* Article 107565. https://doi.org/10.1016/j.chb.2022.107565

Project 3: Grundke, A. (2023a). If machines outperform humans: Status threat evoked by and willingness to interact with sophisticated machines in a work-related context. *Behaviour & Information Technology.* Advance online publication. https://doi.org/10.1080/0144929X.2023.2210688

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

1. Introduction

At a web summit in Lisbon in 2017, Stephen Hawking proposed to “stop for a moment and focus not only on making our AI [artificial intelligence] better and more successful but also on the benefit of humanity” (Koestier, 2017). Over the last decades, scientists and economists have demonstrated their ambition to make machines faster, cheaper, and better as far as it is within the physical possibilities. Even after the limits of Moore’s Law have been reached (Waldrop, 2016), innovative solutions continue to be developed to maintain technical progress. Looking at the closer and more distant past, it becomes evident that creating human-like entities has always fascinated people, vividly illustrated in several artworks, for example, in Ovid’s metamorphoses (ca. 8 C.E./2004), in which the protagonist creates a statue that becomes alive. Other examples of the modern era are Olympia in Hoffmann’s novel “The Sandman” (1816/2012), recent work like the award-winning film “I’m your man” (Schrader, 2021), or the book “Machines like Me” (McEwan, 2019). As such, humans are concerned with the question of how to create similar entities, whether and if so, in what form they can and want to live together with machines, and what effects this could have. In line with Hawking’s demand to have the benefit of artificial intelligence for humanity in mind, computer scientists, engineers, sociologists, and psychologists may strive to determine which technical advances are not only possible but also positive and beneficial for humanity. Following this demand, this research examines whether humans indeed want to interact with sophisticated machines and asks to which point the technological continuation is certainly desired or just an exciting imagination transported by the arts, media, economists, and futurists. This is accompanied by the question under which conditions and in which particular situations humans are willing to accept machines supposedly equipped with their own minds.

So far, having a mind and referring to other minds has been regarded as a crucial and uniquely human ability: The *theory of mind* distinguishes humans from most animals, plants, and inanimate objects, mainly because of the depth and speed humans are capable of demonstrating and adopting this ability (Call & Tomasello, 2008). It describes the process of inferring and reasoning about the perceptions, beliefs, thoughts, or emotions of others and distinguishing them from one's own (Frith & Frith, 2005; Premack & Woodruff, 1978). Humans can not only develop a theory of mind for other humans or animals. They also ascribe mind, awareness, intentionality, justification, and responsibility to non-living beings like robots and artificial intelligence (Banks, 2019; Shank & DeSanti, 2018), suggesting that some kinds of machines, if expressing respective cues, are treated comparably to humans. This becomes exceptionally evident in the *computers as social actors* research (Reeves & Nass, 1996): Deriving social norms humans are familiar with from human-human interaction to various media and also to robots (Eyssel & Hegel, 2012; Kahn et al., 2012; Lee et al., 2006), people attribute gender (Nass et al., 1997), skills (Nass & Moon, 2002), and personality characteristics to machines (Moon & Nass, 1996; Nass & Lee, 2001) and they even expect reciprocity from them (Fogg & Nass, 1997).

Going one step further, people not only automatically perceive a mind in robots expressing social cues but also explicitly relate two components of mind to machines typically associated with human minds: agency and experience (Gray et al., 2007). However, machines supposedly being able to express their minds by these two factors evoked higher aversion than simple technical tools without mental capabilities (e.g., Appel et al., 2020; Taylor et al., 2020). They were rated rather negatively and fell into a so-called *uncanny valley of mind* (Stein & Ohler, 2017; derived from Mori's *uncanny valley*, 1970). The uncanny valley of mind hypothesis states that responses towards human-like machines become more positive until a decrease is observed if entities' mental capabilities become too human-like

(Stein & Ohler, 2017). Yet, the uncanny valley of mind has been considered somewhat in isolation. Most studies in this realm were based on text vignettes and ensured that the phenomenon can, in principle, come about. What has been missing in the scientific landscape so far is embedding a robot's mental capabilities in different concrete scenarios in order to expand the level of knowledge and to be able to make more generalizable statements based on a variety of methods and data, thus motivating the research agenda.

I strive to explore humans' acceptability of robots with sophisticated mental abilities (robots with agency and experience) from various perspectives. Each of the three realized projects highlights one core variable in combination with a concrete scenario. First, *eeriness* evoked by robots with the capability to read another person's thoughts, therefore exceeding humans' capabilities in social interactions (Project 1). Second, *empathy* for robots with mind as a possibility to alleviate the uncanny valley of mind (Project 2). Third and lastly, *status threat* evoked by robots with sophisticated mental abilities and the resulting willingness to interact with such robots in work-related contexts (Project 3).

This three-part work is organized as follows: In my synopsis, I describe the theoretical background of my dissertations' projects and embed the research questions of the respective projects in a larger context. After describing previous research on assessing mental capabilities in machines based on the mind perception dichotomy (Chapters 1.1 and 1.2), I outline the development of artificial intelligence and describe its implementation in robots (Chapter 1.3). In the following, I summarize my research goals resulting from the current state of research (Chapter 1.4). Afterward, I give an overview of the dissertation studies (Chapter 2) by highlighting the three variables forming the core of the respective research project and by describing each project in detail. The second part of this work consists of the three manuscripts (Chapters 3-5). Finally, the General Discussion, as the third part (Chapter

6), summarizes and interprets the findings of my empirical work and offers inspiration for future research.

1.1 Uncanny Valley (of Mind)

More than 50 years ago, Masahiro Mori formulated the uncanny valley hypothesis (Mori, 1970; Mori et al., 2012; for recent reviews, see Diel et al., 2022; Mara et al., 2022). It states that responses to human-like entities such as robots or digital animations become more positive with increasing human likeness until a downturn is observed for highly but not perfectly human-like entities. The feeling of eeriness evoked by human-like-looking machines was revealed to come up at the age of nine years (Brink et al., 2019). In light of the worldwide increase in robot production (International Federation of Robotics, 2022), learning how to handle aversion may also require teaching children to appropriately interact with various forms of technology and media (Nieding & Ohler, 2008).

Traditional uncanny valley research manipulated the human likeness of entities such as robots by adapting the visual appearance (MacDorman & Ishiguro, 2006; Mathur & Reichling, 2009, 2016; Seyama & Nagayama, 2007), while the existence and form of the valley have been discussed controversially over the years (e.g., Bartneck, Kanda, Ishiguro, & Hagita, 2007; Poliakoff et al., 2013). Other research focused on individual differences, context factors, and more functional attributes of the robot to explain negative responses to modern machines (e.g., Lischetzke et al., 2017; MacDorman & Entezari, 2015; Rosenthal-von der Pütten & Weiss, 2015; Tu et al., 2020). Additionally, scholars concentrated on the influence of the perception of a human-like mind in a machine on user acceptance. Stein and Ohler (2017) proposed the uncanny valley *of mind* to summarize responses towards entities with human-like mental capabilities: Following Mori's (1970) approach, the hypothesis describes that aversion is evoked by machines from a certain degree of human-like mental capabilities, while the exit of the uncanny valley of mind so far has mainly remained open (as is also the status of traditional uncanny valley research on appearance, see Mara et al., 2022). Evidence for the uncanny valley of mind was revealed by initial research using a virtual reality chat program to let participants interact with human-controlled avatars or computer-controlled agents. In their work, Stein and Ohler (2017) also manipulated the level of alleged autonomy (scripted vs. self-directed), using an emotional and empathetic dialogue. Eeriness was highest for autonomous artificial agents, delivering positive evidence for the uncanny valley of mind. Apart from virtual agents, robots were also classified to evoke aversion according to humans' mind attribution towards the machines, finding positive evidence for the uncanny valley of mind from a certain degree of mental capabilities (Appel et al., 2020; Gray & Wegner, 2012; Hegel et al., 2008; Wegner & Gray, 2016).

1.2 The Big Two: Agency and Experience

As an underlying framework for the exploration of mind in machines, the mind perception dichotomy (Gray et al., 2007) has gained much scientific attention. Participants were asked to describe the extent to which different types of people, animals, God, and a robot possessed specific mental capacities. A principal component factor analysis revealed that mental capacities could be categorized into *agency* (i.e., self-control, morality, memory, planning, communication, and thought) and *experience* (i.e., the ability to feel emotions and to have a personality). These two factors were named the *Big Two* (for an overview, see Abele & Wojciszke, 2014). This differentiation of mind appeared in psychological research under many different names which share a unifying core, for example, agency versus communion, intellectually versus socially good-bad, masculinity versus femininity, instrumentality versus expressiveness, competence versus morality, dominance versus submissiveness, warmth versus competence, and trust versus autonomy (Abele et al., 2008, 2016; Erikson, 1950; Fiske et al., 2002; Gebauer et al., 2013; Helgeson & Fritz, 1999; Judd et al., 2005; Paulhus & John, 1998; Wiggins, 1979; Wojciszke, 2005; Ybarra et al., 2008). Most of the research in this realm was conducted in Western countries; however, the Big Two appear to be universally valid (Abele et al., 2016; Saucier et al., 2014; Wojciszke & Białobrzeska, 2014; Ybarra et al., 2008). Also objects like robots were automatically categorized within these two dimensions (Gray et al., 2007; Otterbacher & Talias, 2017). Hence, the mind perception dichotomy finds active consideration in human-machine interaction.

Gray and Wegner (2012) built on their prior research (Gray et al., 2007) and investigated the evaluation of entities with several levels of agency or experience in three experiments. After providing correlational hints that a robot with a human-like appearance is attributed experience, they concluded that its uncanniness could be due to violated expectancies, as a supposedly feeling robot violated the expectancy of a robot being an emotionless technical tool. With this, they followed a line of argumentation from the traditional uncanny valley research with a focus on the appearance of machines, in which violated expectancies were also a common explanation of aversion to human-like machines (MacDorman & Ishiguro, 2006). Gray and Wegner's (2012) second experiment presented a super-computer with agency or experience characteristics. The results revealed that experience was predominantly responsible for high ratings of eeriness, which was explained by the perception of experience as a central core of humanness (Bakan, 1966; Gray et al., 2011; Haslam et al., 2005; Knobe & Prinz, 2008). Their third experiment underlined this finding, showing that people without the ability to express experience made humans feel uneasy, which was not the case with the same intensity for a person who lacked agency. The authors concluded that especially experience is part of the fundamental conception of human minds.

Building on Gray and Wegner (2012), Appel et al. (2020) also focused on the mind perception dichotomy. Their research revealed that a robot with experience evoked higher eeriness than a robot with agency, which again evoked higher eeriness than a robot unable to express any form of mind. This result remained stable regardless of the robot's gender and was attenuated by presenting the robot in a nursing context. Moreover, the robustness of this basic finding is accentuated by the fact that this pattern of gradation was not only revealed for robots but also for smart speakers (Taylor et al., 2020) and autonomous cars (Li et al., 2022). In sum, according to the described studies and independent of an entity's embodiment, equipping machines with agency and particularly experience resulted in discomfort and rejection in earlier work.

1.3 Artificial Intelligence in Robots

According to Serholt (2018, p. 252), social robots "are physical, autonomous artifacts that interact and communicate with humans through human social mechanisms, such as natural speech and social cues." For robots to be more than just mechanical shells and meet these requirements, they have to reach an appropriate technical level and the capability to resemble a human-like mind. Along these lines, robots can be categorized according to their capability to simulate human capabilities (Zhang & Lu, 2021). One can distinguish three generations of robots (Liu et al., 2018). The first generation is not equipped with artificial intelligence. Its operations must be represented as step-by-step instructions in its code. A technician must supervise the robot and explicitly start its operations by hand (Lu et al., 2018). The second generation of robots is slightly more sophisticated, the so-called adaptive robots (Zhang & Lu, 2021). These robots are equipped with sensors that can collect simple information, for example, about their working environment and operating objects. An intelligent robot is part of the third generation (Shone et al., 2018). Its sensors are susceptible, and the robot expresses human-like intelligence. Its sensory capabilities go beyond those of humans, and the machine can autonomously adapt to environmental changes and complete complex tasks.

The phenomenon that humanlike properties and characteristics are attributed to non-living beings at all is called anthropomorphism (Epley et al., 2007). The term anthropomorphism is of Greek origin and derives from *anthropos,* meaning "human," and *morphe,* meaning "form" (Duffy, 2003). Possibilities for anthropomorphizing robots vary (Hegel et al., 2011; Złotowski et al., 2015). They range from appearance (Goudey & Bonnin, 2016), the emotions perceived in the machine (Spatola & Wudarczyk, 2021), the intelligence attributed to the machine (Moussawi et al., 2021), and the predictability of its actions (Eyssel et al., 2011), to its verbal (Seeger et al., 2021) and non-verbal (Lugrin et al., 2018) communication. A crucial facet of anthropomorphizing robots is that humans ascribe mind to them (Epley et al., 2007; Waytz et al., 2013). The experiments reported in this thesis mainly concentrate on this facet. One possibility for machines supposedly having human-like mental capabilities is to equip them with artificial intelligence, an envisaged equivalent to let machines resemble human-like minds (Da Xu et al., 2021).

The development of artificial intelligence reaches back many decades. In 1943, artificial neural network research came up with the proposition of the artificial neuron model (Zeng et al., 2012), while the term *artificial intelligence* itself came up at a workshop at Dartmouth College in the United States in 1956 (Nilsson, 2010). After phases of reduced funding and research interest—so-called AI-winters (Floridi, 2020)—the development of artificial intelligence (AI) systems and neural networks experienced a steep surge, partly due to the availability of large publicly usable datasets (Jordan & Mitchell, 2015), for example, the *ImageNet*[1] dataset (Fei-Fei et al., 2021), and the advancements in technological and hardware capacities. While a computer program could defeat the world chess champion Garry Kasparov already in 1997 (Campbell et al., 2002), another noteworthy milestone in the

[1] https://www.image-net.org

development of artificial intelligence was set with the game *Go*. Beginning in 2015, the AI system *AlphaGo* started beating professional Go champions in official matches, which was initially considered to take multiple decades of further research effort (Silver et al., 2016). Since this initial achievement, the AI research organization DeepMind has successfully run AI systems that enable researchers to train more complex systems, learning independently (Silver et al., 2017).

Research and economy are committed to maintaining progress in developing different forms of artificial intelligence (Bughin et al., 2018)—it is estimated that by 2030, AI can deliver an additional global economic activity of 13 trillion USD. Artificial intelligence is a multidisciplinary research field covering computer science, logic, biology, psychology, philosophy, and many other disciplines, as well as the fields of application are diverse. Artificial intelligence, for example, is used to model climate change (Ng, 2022) or to guarantee less biased recruiting procedures (van Esch et al., 2019). As such, artificial intelligence's constant and rapid development has enormously influenced people’s way of living (Huang, Cai, et al., 2019). Consequently, people are directly affected by artificial intelligence’s influence on several affairs and decisions. These days, this influence becomes particularly evident in the discussion about the chatbot ChatGPT based on a series of artificial intelligence models called generative pre-trained transformers (GPT). It is expected to become widely used to write texts in schools, academia, arts, and everyday life to harvest all loads from its users or at least to offer a basic text version upon which users can work (Kelly, 2023; OpenAI, 2022). Other fields of AI applications are speech recognition, image processing, the proving of automatic theorems, and intelligent robots (Duan et al., 2009). By now, there indeed exist systems that can perform tasks that in the past could only be managed and performed by humans (Huang, Huan, et al., 2019), and sophisticated artificial intelligence systems can emulate human brain activities (e.g., Hassabis et al., 2017; LeCun et al., 2015).

The existing and evolving artificial intelligence systems for such tasks can be classified according to their capabilities (Kaplan & Haenlein, 2019): A *weak* artificial intelligence has capabilities that may equal humans in a specific and concrete area. More elaborated is a *strong* artificial intelligence. This type of AI can apply to several areas and equal or even outperform humans in these areas. Lastly, an artificial *super* intelligence applies to any area and can instantly solve problems in various areas and outperform humans. Additionally, this type of artificial intelligence is defined as having self-awareness and is intended to emulate the human mind or even surpass it.

As this dissertation focuses on robots with human-resembling mental capabilities, I now explain how I interpret the term artificial intelligence and describe the level of artificial intelligence with which the robots in my projects are equipped. With the reported classifications in mind, most experiments conducted for this dissertation contrast robots with the capability to resemble a human-like mind in terms of agency and experience with robots that do not have this capability. These so-called tool robots or robots without mind are conceptualized as a combination of the first and second generation of robots, that is, robots that are equipped with sensors but must be programmed by human users, follow their orders, and have no own decision-making ability. If at all, they are equipped with a weak artificial intelligence at most, depending on the respective experiment. On the other hand, sophisticated robots with mental capabilities represent the aforementioned third generation of robots. They are described as equipped with a strong artificial (super-)intelligence that fully resembles or surpasses humans’ mental capabilities. As such, I define sophisticated robots as an embodied form of artificial intelligence in this thesis. These robots are hypothesized to evoke aversion as they fall into the uncanny valley of mind.

1.4 Aims of the Present Research

After explaining essential elements building the foundation for the dissertation, I now outline the individual research goals of my work. The overall research question of my work is how people respond to machines equipped with human-like mental capabilities in diverse social settings and in light of the resulting implications. In my reading, prior work on the mind perception dichotomy (e.g., Appel et al., 2020; Gray & Wegner, 2012; Swiderska & Küster, 2020) has studied the aversion to modern-day machines in a relatively isolated manner, addressing a rather narrow research gap. Scholars primarily focused on vignettes to introduce a robot with sophisticated mental capabilities, presenting new generations of super-machines and entailing that these may be obtainable at some point in the distant future. As such, it is necessary to study the uncanny valley of mind from different perspectives and widen the knowledge in the field by introducing new variables and contexts and using diverse methodologies.

The first goal of this dissertation is to use more diverse methods than text vignettes to study the uncanny valley of mind. A vignette is a "short, carefully constructed description of a person, object, or situation, representing a systematic combination of characteristics" (Atzmüller & Steiner, 2010, p. 128). Since text vignettes are often-used in this research field, internally valid, and a resource-efficient method (Appel et al., 2020; Gray & Wegner, 2012; Shank et al., 2021; Swiderska & Küster, 2020; Ward et al., 2013), some of the experiments presented in this thesis also focus on text vignettes. However, this kind of stimulus may lack ecological validity, as such artificial stimuli and hypothetical scenarios strongly differ from real-world scenarios and are therefore particularly limited in their generalizability outside the respective experimental investigation (Brewer, 2000; Schmuckler, 2001). Therefore, most of the dissertation's experiments use a more varied methodology than text vignettes with the goal to increase the ecological validity over the three projects. I strive to widen the primary knowledge scholars have so far about the uncanny valley of mind, testing its occurrence and replicability for diverse stimuli presented as vignettes, videos, or live interactions. This way, the thesis offers a multi-method approach in which the robot presentations become more and more realistic over the three projects (from text vignettes to videos to a real robot in a live interaction). Importantly, this work does not aspire to go into detail about the traditional uncanny valley with a focus on appearance, even if visuals are implicitly included in some experiments. The definite focus is on the evaluation of machines’ mental capabilities.

The second goal of this dissertation is to have a closer look at boundary conditions of aversion to modern-day machines since responses to them may differ in several situations. In my understanding, more research is needed to explore the mind perception dichotomy and humans’ responses to sophisticated machines in concrete situations. For example, I suggest that embedding a machine with mind in a respective scenario could influence the valence of the machine’s assessment as participants realize why a machine is equipped with mind and understand why it is designed in a specific manner in a specific scenario.

The third goal of this dissertation is to reflect on the possible implications of the uncanny valley of mind and on which additional variables further shape the phenomenon. Since it might be too short-sighted to assume only aversion as a response to sophisticated machines, it is crucial to find out which concomitants of the fact that machines’ mental capabilities can resemble humans’ mental capabilities are decisive for further human-machine interaction.

Taken together, this empirical work extends previous research on the mind perception dichotomy in the uncanny valley of mind literature by (a) using a multi-method approach (text vignettes, videos, interactions), (b) presenting the robot with or without human-like mental capabilities in concrete scenarios to create a realistic experience and (c) interpreting the implications of interactions with robots with mental capabilities with the help of the recorded variables. Along these lines, all studies contribute to expanding the knowledge about humans' assessments of machines with human-like mind under diverse boundary conditions and simultaneously derive recommendations for the further coexistence of humans with robots in specific, concrete social scenarios. Table 1 provides a systematic overview of the research goals concerning the respective project.

Table 1

Research Goals and Bibliographic References

Project	Research goal	Reference
1	Exploring the evaluation of a robot not *expressing* agency or experience but *uncovering* humans' agency and experience	Grundke, A., Stein, J.-P., & Appel, M. (2022a). Mind-reading machines: Distinct user responses to thought-detecting and emotion-detecting robots. *Technology, Mind, and Behavior, 3*(1), 1-12. https://doi.org/10.1037/tmb0000053
2	Exploring human empathy for a harmed robot to investigate a possibility of alleviating the uncanny valley of mind	Grundke, A., Stein, J.-P., & Appel, M. (2023). Improving evaluations of advanced robots by depicting them in harmful situations. *Computers in Human Behavior, 140,* Article 107565. https://doi.org/10.1016/j.chb.2022.107565
3	Exploring the influence of status threat on willingness to interact with a sophisticated machine in a work-related scenario	Grundke, A. (2023a). If machines outperform humans: Status threat evoked by and willingness to interact with sophisticated machines in a work-related context. *Behaviour & Information Technology.* Advance online publication. https://doi.org/10.1080/0144929X.2023.2210688

All projects included in this thesis have been published or are an advance online publication in international peer-reviewed journals (Grundke, 2023a; Grundke et al., 2022a, 2023). Following Open-Science standards, all experiments were pre-registered, and each research project's data, materials, and code can be found on the open science framework (OSF). Besides this, the first project was presented by a poster at the 12th Conference of the Media Psychology Division of the German Society for Psychology in Aachen (Grundke et al., 2021). The second project was presented as a research talk in a high-density session at the 72nd Annual Conference of the International Communication Association in Paris (Grundke et al., 2022b) and as a research talk at the 52nd Congress of the Deutsche Gesellschaft für Psychologie in Hildesheim (Grundke et al., 2022c). The third project was presented as a poster at the 13th Conference of the Media Psychology Division of the German Society for Psychology in Luxembourg in September 2023 (Grundke, 2023b). Before being invited to present my research at the mentioned conferences, all contributions were peer-reviewed.

2. Overview of Dissertation Studies

The evidence described above demonstrates that there is no need to explore the basic fact that people feel an aversion to machines with too human-like mental capabilities—according to the literature, this effect seems relatively stable if text vignettes were used as stimuli. However, various co-contextual variables and extensions of this finding have remained much more open. The theory of mind, for example, implies that humans can infer the mental processes of others (Frith & Frith, 2005). But what would happen if a robot equipped with artificial intelligence was not only capable of doing the same as humans (= developing a theory of mind for humans, which in and of itself can be unexpected) but could even exceed people's capabilities and not only "read" another's emotions but also thoughts? As such, the first project extends research on the theory of mind in combination with uncanny valley of mind aspects: The users' acceptance of robots not being able to *express* their mind (Appel et al., 2020; Gray & Wegner, 2012) but being able to *analyze* the emotions and thoughts of human counterparts is studied (Project 1).

After focusing on this advancement of the mind perception dichotomy and due to the steadily increasing amount of sophisticated machines (Bryndin, 2020; Hildt, 2019; International Federation of Robotics, 2022), the question arose whether there is a possibility to mitigate the uncanny valley of mind. To the best of my knowledge, how this phenomenon can be counteracted has not yet been explored. Can the uncanny valley of mind be alleviated in concrete scenarios? By harming a robot and thus evoking empathy, human empathy for a harmed robot is investigated as a suitable mediator to increase the robot’s likeability. On the other hand, in line with the uncanny valley of mind hypothesis, the likeability of robots with mental capabilities in neutral situations is considered to be lower (Project 2).

Finally, I put my third research question in a relevant scenario for scientists, practitioners, and enterprises. How would sophisticated robots influence the future world of work—do people want to interact with robots that can fulfill several tasks requiring human-like mental capabilities? The third project explores how status threat (Pettit et al., 2013) evoked by a machine with human-like mental capabilities affects the future willingness to interact with such a machine in a work-relevant scenario, thus offering research with a high potential for application. Additionally, an important goal is to widen the realism of the methods used by offering interactions with a robot and a simulated artificial intelligence without embodiment (Project 3). After the first project uses text vignettes, the second project uses videos, and the third project offers interactions. This way, I increase the realism of robot presentations and thereby the ecological validity over the three projects step by step. Table 2 shows a methodological comparison of the three projects. In the following, I will explain the selection of core variables (Chapter 2.1) and describe the three projects (Chapters 2.2.-2.4).

Table 2

Methodological Comparison of Projects Presented in the Thesis

Project	Sample size	Method	Outcome measures
1	N_1 = 335 N_2 = 536	▪ Text vignettes	▪ Eeriness (Gray & Wegner, 2012) ▪ Concerns about human identity (Kamide et al., 2012, two items; Stein et al., 2019, three items) ▪ General evaluation (Appel et al., 2019) ▪ HEXACO-60 questionnaire (Ashton & Lee, 2009)
2	N_1 = 559 N_2 = 396	▪ Text vignettes ▪ Videos ▪ Videos with vocal explanations	▪ Likeability (Bartneck et al., 2009) ▪ Empathy (Oswald, 1996)
3	N_1 = 104 N_2 = 589	▪ Manipulated performance feedback ▪ Live interaction with the robot NAO ▪ Live interaction with AI	▪ Status threat (Pettit et al., 2013) ▪ Willingness to interact (Robinson et al., 2018) ▪ Objective performance ▪ Perceived usefulness (Davis, 1989) ▪ Mindset about human minds (Dang & Liu, 2022a)

2.1 Selection of Core Variables

In Mori's (1970) initial work on several replicas with increasing human likeness (independent variable), the dependent variable was labeled with the Japanese neologism *shinwakan*, a combination of social presence and connection, translated to English as *affinity* (Mori et al., 2012). According to Diel et al. (2022), a common approach to assess this evaluation in the traditional uncanny valley research is using related constructs like *likability*, *aesthetics*, *familiarity*, and reverse-scaled *threat* and *eeriness.* Recent meta-analytic evidence (Mara et al., 2022) highlighted that, over the years, uncanny valley research mainly referred to the human likeness and the likeability dimension of the Godspeed

Questionnaire (Bartneck et al., 2009) as dependent variables to assess people's evaluation of machines with human-like appearances. In contrast, the uncanny valley indices of Ho and MacDorman (2010, 2017) were barely used for measuring people's evaluation of machines with human-like appearances.

In the newer uncanny valley of mind research, the amount of explored dependent variables is even smaller, mainly focusing on eeriness (Gray & Wegner, 2012) and kinds of threat to human uniqueness (Stein et al., 2019; Złotowski et al., 2017). These variables are connotated negatively, which aligns with the core of the uncanny valley (of mind) assumption. As these two variables received much attention in the research field, I also concentrate on eeriness and a form of threat as crucial variables to measure aversion to sophisticated machines (Projects 1 and 3). Since they are elucidated from a new perspective, I contribute advanced insights to the state of knowledge around these variables.

On the other hand, research also highlights the positive consequences of equipping a machine with human-like mental capabilities, such as feeling empathy for them (Choi et al., 2021; Nijssen et al., 2019). Empathy can be evoked by presenting robots in situations in which they are harmed (Menne & Schwab, 2018; Rosenthal-von der Pütten et al., 2013). Combining this evidence and intending to find a way out of the uncanny valley of mind, human empathy for a harmed machine is explored as a positively connotated variable (Project 2). In the following three subchapters about the described projects, a theoretical overview of the core variable used in each project is given, followed by a description of the conducted experiments.

2.2 Reversing the Uncanny Valley of Mind Perspective (Project 1)

2.2.1 Core Variable: Eeriness

The variable eeriness is used both in the uncanny valley (Ho & MacDorman, 2017) and the uncanny valley of mind (Appel et al., 2020) research. The synonymous term "the uncanny" became famous by the German psychoanalyst Sigmund Freud in 1919 when he answered an earlier publication by Ernst Jentsch (1906/1977). The English term "uncanny" is rooted in Northern and Scots dialects ("beyond ken/knowledge") and became connotated with strangeness, untrustworthiness, and the supernatural in the 18th century. The German term "unheimlich" implies a departure from home and familiar qualities (Salvesen, 2021; Vidler, 1992). As a reason for this feeling, Jentsch (1906/1977) posited human doubts on whether a living being indeed is an animate being or whether an inanimate being may actually be alive. Freud (1919/2020) expanded this explanation by assigning the unfamiliar to eeriness and the familiar, consequently focusing on eeriness by ambiguity. Meanwhile, it has been meta-analytically shown that eeriness is a well-established variable to operationalize robot acceptance (Diel et al., 2022).

The uncanny valley researcher Karl MacDorman (2006) described eeriness as a response to a mismatch between human expectations and a robot's behavior. Also, Weis and Wiese (2017) defined uncanniness as a result of a cognitive conflict due to different categories being activated simultaneously. Another approach is to distinguish uncanniness by a bidirectional spine-tingling subfactor (uninspiring–spine-tingling, boring–shocking, predictable–thrilling, bland–uncanny, and unemotional–hair-raising) and a bidirectional eerie subfactor (dull-freaky, predictable-eerie, plain-weird, ordinary-supernatural; Ho & MacDorman, 2017). Benjamin and Heine (2023) criticized these uncanny valley indices for assessing people's judgments of visual stimuli instead of the feeling of uncanniness as an affective experience. Moreover, since these approaches were established to evaluate the appearance of human-like robots, researchers in the uncanny valley of mind field needed to develop own scales to assess eeriness as a response to human-like mental capabilities. Items consisting of one word like "uneasy," "unnerved," or "creeped out" were used by Gray and Wegner (2012) to assess participants' feelings of eeriness evoked by machines with mental capabilities. Other work on the perception of robot mind (Appel et al., 2020) or creative artificial intelligence (Messingschlager & Appel, 2022) also relied on this scale. Therefore, this measurement of the eeriness variable is used as the primary dependent variable in the first publication included in this thesis.

2.2.2 Description of Project 1

Human aversion evoked by machines that can express mind in forms of agency or experience has been part of numerous research (Appel et al., 2020; Gray & Wegner, 2012; Li et al., 2022; Taylor et al., 2020). However, one important aspect when thinking about mind is not only being able to express it or being aware of one's own mind (Adolphs, 2009) but particularly about processing the mental states of others (Premack & Woodruff, 1978). An elaborated theory of mind is an important human ability (Adolphs, 2009). Therefore, an interesting question is how people evaluate robots equipped with a sophisticated artificial intelligence so that the machine possesses comparable mental and, thus, no longer solely human characteristics.

To answer this question, the first project focuses on the perspective of a robot not explicitly able to *express* its own mind but being able to *read* human minds. Using a between-subjects design, participants are informed about the innovative robot Ellix as capable of analyzing people's (a) emotions or (b) thoughts with the help of sensors and artificial intelligence. A tool robot without mind-reading capabilities is presented (c) as a control group in Experiment 1. While a robot expressing experience was evaluated to be eerier than a robot expressing agency in earlier work, the opposite gradation with the reversed perspective is assumed. The hypothesis states that reading humans' thoughts (agency) would be eerier than reading humans' emotions (experience) due to what people are familiar with from human-human interaction. Apart from shifting the perspective taken in the prior uncanny valley of mind research, a crucial cognitive component of recognizing agency capabilities (i.e., thoughts) is introduced, as earlier work in this area (Kang & Sundar, 2019; Stein & Ohler, 2017) mainly focused on machines recognizing another's experience capabilities (i.e., studying entities that simulate to be capable of empathizing).

The study uses text vignettes to present a robot supposedly being able to unfold humans' agency or experience with the help of a sophisticated artificial intelligence system. Text vignettes are presented to ensure the credibility of the robots' alleged capabilities to read emotions or thoughts. This procedure is expected to be a suitable approach to entering new research fields (Aguinis & Bradley, 2014). In human-robot interaction (e.g., Appel et al., 2020; Gray & Wegner, 2012), for example, vignettes are used to study reactions to machines with capabilities not yet implemented in them due to technological limitations, like is currently the case for thought detection.

To get an overall impression of people's assessment of machines with mind-reading capabilities, eeriness, concerns about human identity, and the general evaluation of the robot serve as dependent variables in Experiment 1. A second online experiment, additionally considering the HEXACO personality dimensions (honesty-humility, emotionality, extraversion, agreeableness, conscientiousness, openness to experience) is conducted to draw conclusions about the generalizability of the findings revealed in Experiment 1.

2.3 Alleviating the Uncanny Valley of Mind (Project 2)

2.3.1 Core Variable: Empathy

As there are numerous definitions of empathy, it is crucial to clarify in which sense I understand the often-researched variable. Eklund and Meranius (2021) provided a consensus definition of empathy by conducting a review of reviews. They summarized that for the occurrence of empathy, an empathizer must (1) understand, (2) feel, and (3) share another person's world by making (4) a self-other differentiation.

People do not only empathize with persons but also empathize with robots (Hoenen et al., 2016; Rosenthal-von der Pütten et al., 2013; Seo et al., 2015). There were reports of people reacting with empathy for robots, for example, for the hitchhiking humanoid robot Hitchbot: On a tour of the United States, the robot was beaten up by people, which caused great empathy for the robot and a lack of understanding for the perpetrators (Darling, 2015; Gunkel, 2018). Furthermore, the robot company Boston Dynamics received negative criticism for its attempts to unbalance robots by hitting them to test the technology and mechanics (see Küster et al., 2020, for an overview of comments), implying that people felt empathy for a—in their perception—teased robot.

As soon as empathy is evoked, the question arises as to the consequences. Whether empathy positively or negatively influences robots' evaluations has been part of theoretical considerations, especially in combination with the robot's appearance. For example, it was postulated that a human's "complete" feeling of empathy for a humanoid entity might be inhibited because of its perceived visual imperfections (Misselhorn, 2009). Scholars assumed that empathy for robots with a human-like appearance proceeded in the same form as the uncanny valley curve and was therefore discussed as an explaining factor of eeriness (MacDorman et al., 2013; MacDorman & Entezari, 2015; Misselhorn, 2009). Hence, Vanman and Kappas (2019) observed a paradox: Because of a robot's human likeness, people feel empathy for it, and (at the same time) they feel threatened by it. However, empirical research (Diel & MacDorman, 2021) could not confirm that empathy may be an antecedent of aversion and suggested empathy, which on the contrary, could even be induced by human likeness (de Jong et al., 2021; Riek et al., 2009), to be a promising chance to prevent negative evaluations of human-like robots.

One step further, humans showed similar brain activities when empathizing with humans and robots (Gazzola et al., 2007; Rosenthal-von der Pütten et al., 2013, 2014).

Regarding a robot-robot comparison, higher brain activity in response to human-like-looking robots was measured compared to relatively simple robots (Krach et al., 2008). Humans also attributed more intelligence to human-like robots than to non-human robots or computers (Hegel et al., 2008). This described evidence was primarily obtained concerning machines' human-like appearances. Nevertheless, the empirical evidence already hints that similarities between machines and humans can lead to higher empathy for the respective technology, and findings from interpersonal literature show that empathy is linked with likeability (Brooks et al., 2014; Johnson et al., 1983; Meuwese et al., 2017). These results seem encouraging to explore empathy in response to machines expressing mental human likeness as a possible solution to alleviating aversion to modern-day machines.

By now, research has not explicitly focused on the influence of empathy in combination with robot mind on robot evaluation—a crucial research gap that the second project addresses. So far, it was revealed that empathy turned out to be exceptionally high for robots described as having a mind: the perception of a tortured entity being able to *experience* the torture was fundamental for empathy to emerge (Choi et al., 2021; Nijssen et al., 2019). To ensure that empathy is evoked, the experiments described in the following section build on prior studies in which robots were presented in harmful situations to evoke human empathy for these machines (Menne & Schwab, 2018; Rosenthal-von der Pütten et al., 2013).

2.3.2 Description of Project 2

Based on evidence about empathy as a possibility to bond individuals (Batson & Ahmad, 2009; Bruneau & Saxe, 2012; Malhotra & Liyanage, 2005; Paluck, 2009; Plutchik, 1987), the second study's goal is to investigate people's empathy for a robot as a possibility to alleviate the uncanny valley of mind. The robot is presented in a situation in which it is intentionally harmed to induce human empathy (Brščić et al., 2015; Menne & Schwab, 2018; Rosenthal-von der Pütten et al., 2013). As such, the situation in which the robot is presented (neutral vs. harmful) and the robot's mind (with mind vs. without mind) is manipulated. The two experiments follow a 2×2 between-subjects design. The underlying idea is that a robot equipped with mental capabilities in a neutral situation evokes low likeability, hypothesizing to fall into the uncanny valley of mind. On the other hand, a robot with mental capabilities in a harmful situation should evoke strong empathy and, finally, likeability (Choi et al., 2021; Lucas et al., 2016; Nijssen et al., 2019; Yam et al., 2020). Mind and situation serve as independent variables, empathy serves as the mediator variable, and likeability serves as the dependent variable. An interaction effect of situation and robot mind on likeability is hypothesized, and a moderated mediation model is formulated.

In the first experiment, text vignettes are used to manipulate the robot's mind, followed by a video showing the robot named Atlas (Boston Dynamics), in which the robot is either harmed by a human or not. In the second experiment, the two manipulations are combined to exclude the possibility that participants of the first experiment only focused on visuals or the text vignette when evaluating the robot but did not perceive the two manipulations as a unit. Therefore, I created four manipulated videos, introducing the robot Atlas as a robot with or without mind and either being harmed or not. A female voice-over with subtitles explained the robot's mental capabilities. This way, it is explored whether aversion to sophisticated machines comes up when solely videos are used as stimuli in which situation and mental capabilities are manipulated at the same time.

2.4 Interacting With Sophisticated Machines in a Work Scenario (Project 3)

2.4.1 Core Variable: (Status) Threat

As for empathy, interpersonal literature seems to be a fruitful starting point for thinking about people's responses to modern-day machines in the realm of threat. Based on previous interpersonal research (Riek et al., 2006; Stephan et al., 1999), Złotowski et al. (2017) distinguished realistic and identity threats that robots can evoke. Assuming that humans perceive robots as part of an outgroup, robots can pose a threat to human identity and uniqueness on the one hand and a realistic threat on the other. The latter describes threats in the realm of resources, human safety, human jobs, and well-being, while the former is related to an ingroup's uniqueness, values, and distinctiveness. As the terms *realistic threat* and *identity threat* may sometimes interfere, Stein et al. (2019) proposed to use the term *threat proximity*, differing realistic and distal threats. These references demonstrate that robots can threaten humans on a symbolic level (identity, distal) and on a practical level (realistic, proximal). Particularly the symbolic threat could be reinforced by the blurred boundaries between humans and sophisticated machines with human-like mental capabilities. Therefore, a challenged human perception of one's own identity as part of an elaborate human species (Fuller, 2014; Kaplan, 2004; MacDorman, Vasudevan, & Ho, 2009; Schultz et al., 2000) served as an explanation of the emergence of threat in the uncanny valley (of mind) literature (e.g., MacDorman & Entezari, 2015; Stein & Ohler, 2017; Stein et al., 2019).

Regarding the increasing number and fast-developing quality of systems with an elaborate mind based on an artificial intelligence (Lu et al., 2019; Pelau et al., 2021), I scrutinize the so-far investigated construct of threat in human-robot interaction from another, new perspective embedded in a practical scenario. In the case of threat, I present a threatening machine in a work-relevant scenario. Apart from its practical relevance, the work-related scenario is chosen because the feeling of being threatened by other co-workers can be quite common, and a competition in essential domains influences humans' perception of their rank, potentially leading to a feeling of threat (Campbell et al., 2017; Cohen-Charash, 2009; Duffy et al., 2012; Dunn & Schweitzer, 2006; Smith & Kim, 2007; Tesser, 1988).

Building on this interpersonal literature, I draw parallels between the two variables threat to human uniqueness (Stein et al., 2019) and status threat in workplace contexts (Reh et al., 2018). The former was used in the uncanny valley of mind research, and the latter was studied in workplace contexts. In my understanding, if sophisticated robots challenge human uniqueness by diminished differences in mental capabilities, they simultaneously challenge humans for their (e.g., professional) rank and thus threaten their status. Similarly, human status at work could be devalued using human-like machines and therefore threaten both on the symbolic and practical levels. As such, I introduce the variable status threat, so far used in interpersonal workplace literature, to human-robot interaction.

2.4.2 Description of Project 3

While the first two projects are based on vignettes and videos, this project uses a different method. It offers an interaction with the humanoid robot NAO (Aldebaran) that I operate myself (Experiment 1) and a simulated interaction with an artificial intelligence (Experiment 2) with the goal to connect the findings of the prior projects in a larger context. Since the number of human-robot collaborations in upcoming work environments has been considered to increase (D'Cruz & Noronha, 2021), I explore a joint task execution between a sophisticated robot and a study participant and follow the perspective that human-robot *collaborations* will become common-spread in the following years, as opposed to cases where humans are totally replaced by robots (Woo, 2020). A machine and a study participant are presented as collaborators on verbal-creative tasks to investigate the potential consequences of upcoming status threat in response to the machine's excellent performance in such a work-relevant scenario. Although machine and human are presented as team members facing the challenge together, it is pointed out that the participant's individual endeavors in the task are highly crucial. Participants are told to be preferentially treated in a money raffle if they contribute more to solving the task than their robotic colleague. As such, I build on the evidence that threat was highest for internal and not for external rivalry (Menon et al., 2006).

The participant and the machine are asked to fulfill word-matching tasks, meant to require agency and experience to solve. In the first experiment, conducted as a laboratory study (via Zoom), these tasks are taken from the IST-2000R intelligence test (Liepmann et al., 2007). In the follow-up online study, the material is self-created and inspired by the material of the first experiment. Participants perform five rounds with eight tasks each, leading to a total amount of 40 tasks. After each round, manipulated performance feedback is offered and serves as the independent variable (main contributor to the task: human vs. machine). After these tasks, humans assess their feeling of being status-threatened by the machine. Status threat is used as a mediator variable derived from the variable threat to human uniqueness in other studies on human-machine interaction, but precisely with a focus on the work context. This way, it is explored whether people feel status-threatened by machines presented to be co-workers, investigating the threat variable from a new perspective.

As dependent variables in Experiment 1, I consider the participant's objective performance and willingness to interact with the robot NAO. In the second experiment, I only consider the willingness to interact with the artificial intelligence as the main dependent variable and use an artificial intelligence without embodiment to exclude potential confounding variables due to NAO's childlike anthropomorphic design (Laban et al., 2021; Rosenthal-von der Pütten & Krämer, 2014; Wu et al., 2012; Zhang et al., 2016). Importantly, it is examined whether the machine's perceived usefulness (Technology Acceptance Model; Davis, 1989) predicts status threat evoked by and willingness to interact with the machine. Furthermore, the participants' mindset about human minds is considered to explain the willingness to interact with the sophisticated machine (Dang & Liu, 2022a).

3. Project 1 | Mind-Reading Machines

Distinct User Responses to Thought-Detecting and Emotion-Detecting Robots

Andrea Grundke

Jan-Philipp Stein

Markus Appel

Status:

Published in the journal *Technology, Mind, and Behavior*

Formal Citation/Reference:

Grundke, A., Stein, J.-P., & Appel, M. (2022a). Mind-reading machines: Distinct user responses to thought-detecting and emotion-detecting robots. *Technology, Mind, and Behavior, 3*(1), 1-12. https://doi.org/10.1037/tmb0000053

Abstract

Human-like robots and other systems with artificial intelligence are increasingly capable of recognizing and interpreting the mental processes of their human users. The present research examines how people evaluate these seemingly mind-reading machines based on the well-established distinction of human mind into agency (i.e., thoughts and plans) and experience (i.e., emotions and desires). Theory and research that applied this distinction to human-robot-interaction showed that machines with experience were accepted less and were perceived to be eerier than those with agency. Considering that humans are not yet used to having their thoughts read by other entities and might feel uneasy about this notion, we proposed that thought-detecting robots are perceived to be eerier and are generally evaluated more negatively than emotion-detecting robots. Across two pre-registered experiments (N_1 = 335, N_2 = 536) based on text vignettes about different kinds of mind-detecting robots, we find support for our hypothesis. Furthermore, the observed effect remained independent of the six HEXACO personality dimensions, except for an unexpected interaction with conscientiousness. Implications and directions for future research are discussed.

Keywords: uncanny valley, mind perception, detector robots, personality, human-robot interaction

3.1 Introduction

Thoughts are free, who can guess them?
They fly by like nocturnal shadows.
No person can know them, no hunter can shoot them
with powder and lead: Thoughts are free!

First verse of the German folk song *The thoughts are free [Die Gedanken sind frei]*

Since antiquity, humans have found relief in knowing that our cognitions cannot be accessed by anyone but ourselves (e.g., Cicero, ca. 52 B.C.E./1977). Due to the constantly advancing development of artificial intelligence, however, this freedom of thoughts (as expressed in the German folk song *Die Gedanken sind frei*) is in peril. Likewise, artificial intelligence is increasingly used to evaluate human emotions. How do humans respond to these mind-reading technologies?

Human (and non-human) mind can be distinguished into agency (thoughts and plans) and experience (emotions and desires, Gray et al., 2007), a distinction that has recently been applied to human-machine-interaction (Appel et al., 2020; Gray & Wegner, 2012; Taylor et al., 2020). The respective studies show that machines with experience are less well-accepted and often perceived to be eerier than those with agency. Yet, it remains unclear how people react to robots who do not express their own mental states but instead detect the mind of the human user. In two pre-registered experiments, we apply the agency–experience distinction to juxtapose robots that can detect thoughts (thought detectors) with those that can detect emotions (emotion detectors).

Contrary to the effects for self-expressing machines, we propose an opposite effect for mind detection: Thought-detecting robots are expected to be eerier than emotion-detecting robots. Additionally, our second experiment applies the HEXACO model of personality in order to examine whether individual differences moderate this effect.

Humanoid Robots and the Uncanny Valley

The production and diversification of service robots is on the rise. The COVID-19 pandemic led to an increased demand for cleaning and disinfection robots, food and medication delivery robots, and edutainment and interaction robots (International Federation of Robotics, 2020). At the same time, a multi-wave international study showed that attitudes towards robots have become more negative over the last years (Gnambs & Appel, 2019). Faced with observations such as these, people may turn to scientific evidence to look for explanations.

A popular framework underlying negative responses to robots is the *uncanny valley* model (Mori, 1970; Mori et al., 2012; for reviews see Kätsyri et al., 2015; Wang et al., 2015; Złotowski et al., 2015). It states that responses to human-like entities such as robots or digital animations get more positive with increasing human likeness until a steep drop is observed for highly (but not perfectly) human-like entities. Whereas traditional uncanny valley research manipulated the human likeness of entities such as robots by changing their visual appearance (MacDorman & Ishiguro, 2006; Mathur & Reichling, 2009, 2016; Seyama & Nagayama, 2007), more recent research focused on functional features of the respective technologies, as well as user variables and context factors (e.g., Broadbent, 2017; Lischetzke et al., 2017; MacDorman & Entezari, 2015; Mara & Appel, 2015; Piwek et al., 2014; Rosenthal-von der Pütten & Weiss, 2015; Tu et al., 2020). Also, adhering to a psychological viewpoint rather than merely focusing on visuals, the ascribed mind of robots could be a key to understanding negative responses to robots (e.g., Gray et al., 2007).

Ascribing Mind to Machines

Theory and research suggest that negative responses to human-like robots may depend strongly on the perception of a human-like mind in a machine (Gray & Wegner, 2012; Hegel et al., 2008; Stein & Ohler, 2017; Wegner & Gray, 2016). Indeed, at the age of nine, children already classify robots as more or less scary depending on whether they attribute a human-like mind to them (Brink et al., 2019).

As an underlying framework for this line of research, the mind perception dichotomy by Gray et al. (2007) has gained a lot of attention in recent years. In their initial research, Gray and colleagues asked participants to describe the extent to which different types of people, animals, God, and a robot possessed specific mental capacities. Based on these data, a principal component factor analysis revealed that mental capacities might be categorized into *experience* (i.e., the ability to feel emotions, have a personality and a consciousness) and *agency* (i.e., self-control, morality, memory, planning, communication, and thought). According to further research, it is especially experience that seems to be a fundamental part of how people conceptualize the *human* mind and therefore humanness in general (Gray et al., 2011; Haslam et al., 2005; Knobe & Prinz, 2008).

Considering this paradigm, as well as some alternative theoretical approaches (e.g., Malle, 2019; Weisman et al., 2017), the notion of mind perception has become increasingly relevant in the field of human–robot interaction. For instance, Gray and Wegner (2012) combined the uncanny valley hypothesis with the mind perception dichotomy and showed that machines equipped with experience were rated as much more discomforting and uncannier than those demonstrating agency. In a similar vein, it has been shown that participants rather assigned agency characteristics than experience characteristics to robots (Brink et al., 2019; Gray et al., 2007; Wegner & Gray, 2016). Further building upon the work by Gray and Wegner (2012), Appel and colleagues (2020) presented evidence that a robot with experience was perceived to be eerier than a robot with agency, followed by a robot who merely served as a tool. Indicating notable generalizability, this finding was conceptually replicated for smart speakers in a recent study (Taylor et al., 2020).

Mind Detection by Machines

The mind perception literature has profoundly advanced the scholarly understanding of how people evaluate autonomous technology. However, we note that the scholarly interest in this regard has mainly revolved around the perception of (artificial) minds in machines—yet hardly looked at the other direction, that is, user evaluations of machines analyzing the human mind. Arguably, while this idea might have been dismissed as technically impossible a couple of decades ago, recent technological advancements have turned mind detection by robots into an imminent reality.

By now, advanced software that allows social robots and other technical devices to recognize the emotions of human users can reach impressive levels of accuracy (e.g., Affectiva, 2018; Alonso-Martin et al., 2013; Chen et al., 2020; Microsoft Azure, 2018), leading to an increased scientific interest in digital forms of emotional recognition and mind perception (Banks, 2019; Bianco & Ognibene, 2019; Dissing & Bolander, 2020; Gray & Wegner, 2012; Kang & Sundar, 2019; Stein et al., 2020). Along these lines, it has been suggested that machines might even become able to detect not only human emotions but also human thoughts in the future—a feat that would reach clearly beyond the capabilities of their human creators. In fact, current-day technology already heralds the rise of these possibilities, as machines have been able to deduce internal thought from eye movements (Huang, Li, et al., 2019), create their own theory of mind for humans via computational models (Breazeal et al., 2009; Brooks & Szafir, 2019; Dissing & Bolander, 2020), or use language processing to identify political views (Colleoni et al., 2014), and suicidal intentions (Walsh et al., 2018).

At the same time, it remains unclear how people react to these emerging technologies. Human behavior, appearance, and skills are often used as a reference point when designing modern-day technology (e.g., Eyssel et al., 2012; Huang & Mutlu, 2013; Niculescu et al., 2013; Salem et al., 2011), but users do not always appreciate impressions of humanness in their machines. Indeed, several studies showed that once new technologies threaten human uniqueness, they are typically met with strong aversion (e.g., Müller et al., 2020; Złotowski et al., 2017). Even more so, social cognitive abilities such as mind-reading might play a particular role in this regard (Stein & Ohler, 2017), as our ability to infer and analyze the emotions of those around us has long served as a distinct advantage to our species (Darwin, 1872/2009; Nesse, 1990). Considering this fear of losing our distinctiveness to machines, it appears likely that people might be wary of robots that detect others' emotions—or even surpass this ability with the possibility to "read" cognitions as well.

To this day, however, only a few psychological studies have actually examined user responses to mind-detecting technology in an empirical manner. Kang and Sundar (2019) found that a robot was evaluated more negatively if it correctly interpreted humans' sarcasm than if it failed to recognize this aspect of human behavior. Similarly, research by Stein and colleagues (2019) suggested that an artificial intelligence capable of analyzing participants' personality traits might be seen as threatening. Yet, previous efforts such as these were clearly limited by the fact that they either focused only on emotional aspects of mind or kept the scope of the detection abilities ambiguous (e.g., Kang & Sundar, 2019; Stein et al., 2019; Stein & Ohler, 2017). Therefore, a structured exploration of user reactions to distinct forms of mind detection by machines is all but needed to close an important research gap in the field of human–computer interaction.

The Current Research

We assumed that—unlike the previously documented responses to robotic agency versus experience (e.g., Appel et al., 2020; Gray & Wegner, 2012)—user evaluations might turn out quite differently for the *detection* of *human* agency versus experience by social robots. More specifically, we expected a reversed effect: A robot's ability to analyze human experience should be perceived as less threatening and less uncanny than a robot's ability to analyze users' agency.

In their daily life, humans are generally quite used to other communicators detecting their emotions (Darwin, 1872/2009; Nesse, 1990), whereas precise thought detection is an ability largely unknown from the realm of human-to-human interaction. In turn, people are used to controlling their emotional displays and they have learned to deal with the unintentional communication of emotions (Tamir, 2016), yet they are much less experienced in controlling their thoughts or in coping with the unintentional communication of thoughts and plans. To illustrate this argument, one may consider the embarrassment that people tend to experience when human communication partners detect and interpret a Freudian slip, revealing supposedly true yet hidden thoughts and plans. Based on the large number of studies that have emphasized perceived control as a fundamental prerequisite of positive human-machine interactions (Kang, 2009; Roubroeks et al., 2010; Stein et al., 2019; Sundar, 2020; Zafari & Koeszegi, 2020; Złotowski et al., 2017), we therefore expected a clear advantage of emotion-detecting over thought-detecting machines in participants' evaluations.

Apart from our main outcome variable eeriness (Gray & Wegner, 2012), which remains one of the most well-established ways of operationalizing robot acceptance (Diel et al., 2022), we used two additional dependent variables to get a more general overview of participants' assessment of this type of robotic technology. First, we focused on concerns about human identity, which emerged as a meaningful predictor of technology-related experience in previous research (Stein et al., 2019). More specifically, this variable assesses the extent to which users consider a machine as a symbolic threat to the distinctiveness of the human species (i.e., their uniquely human identity)—an impression that has, in turn, been linked to the unwillingness to further interact with technology (e.g., Kang & Sundar, 2019; Stein et al., 2019; Złotowski et al., 2017). As we presented emotion detectors (which have the same abilities as humans) and thought detectors, whose capabilities even exceed those of humans, we assumed that traditional human–machine boundaries could become blurred, resulting in a meaningful effect expressed by this variable. Second, the general evaluation of the new technology was assessed (Appel et al., 2019), in order to observe reactions towards the presented robots in a more generalizable way.

To implement the desired manipulation of robot characteristics, we used vignette texts—as previous work in the field of mind perception (Appel et al., 2020; Gray & Wegner, 2012; Swiderska & Küster, 2020; Ward et al., 2013) showed that this method can be an internally valid and efficient means to convey specific technological possibilities. In our first experiment, descriptions of an innovative robot able to analyze humans' agency or to analyze humans' experience were presented. As a control group, we presented a description of a robot who merely served as a tool without any sophisticated analysis abilities. Based on the theory and research outlined above, the following hypotheses guided Experiment 1:

H1: The thought detector robot will evoke higher eeriness than the emotion detector robot (H1a), whereas the robot without analysis abilities will evoke the least eeriness (H1b).

H2: The thought detector robot will evoke stronger concerns about human identity than the emotion detector robot (H2a), whereas the robot without analysis abilities will evoke the least concerns (H2b).

H3: The thought detector robot will yield a more negative general evaluation than the emotion detector robot (H3a), whereas the robot without analysis abilities will yield the most positive general evaluation (H3b).

In addition to providing a replication of the effects tested in Experiment 1 (by using the same vignette texts), the second experiment examined the influence of users' individual differences on the acceptance of detector robots using the well-established HEXACO model of personality (Ashton et al., 2004). The hypotheses addressing the role of the users' personality will be introduced after the discussion of Experiment 1. Both experiments were pre-registered, with changes in the hypothesis numbering and exclusion criteria being documented in the supplement. The pre-registrations, data, codes, and an online supplement can be found at https://doi.org/10.17605/OSF.IO/U52KM.

3.2 Experiment 1

Method

Participants

A power analysis with G*Power (Faul et al., 2007) recommended at least 200 participants assuming a small to medium effect size of f = .20 (with alpha-error probability = .05, and power = .80) for the two-group fixed effect expected in Hypothesis 1a. Another 100 participants constituted the control condition, resulting in 300 participants. We invited 450 U.S.-American residents from the MTurk online participant pool (hit approval rate > 97%, hits > 1000), in order to have a buffer if careless responding occurred. Of the 443 completions, 44 participants did not have sufficient English skills, as indicated by two control questions, and were therefore not included in our statistical analyses (Kennedy et al., 2020). One additional participant failed an included attention check item and another three participants had large (> ±3 years) deviations when asked twice about their age. Moreover, 21 participants were excluded because their participation time was lower than 100 seconds (*n* =

4) or higher than 920 seconds (*n* = 17). Another 39 participants interchanged the thought detector robot and the experience detector robot in the manipulation check and were excluded (see supplement for additional information). As such, the final sample consisted of 335 participants (154 female, 176 male, 5 non-binary or no answer) with an average age of 39.33 years (*SD* = 12.00, ranging from 21 to 75 years). Exploratory analyses revealed that age and gender did not moderate the influence of the robot manipulation on the dependent variables (see additional analyses on gender and age for both experiments in the supplement).

Procedure

We asked participants to give informed consent before starting the online experiment. Following their random assignment to one of the three conditions, participants were presented with the respective vignette text matching their group. Subsequently, we asked them to fill in the chosen user evaluation questionnaires. Sociodemographic information and questions to identify careless responding and low English proficiency followed (Kennedy et al., 2020; Meade & Craig, 2012; see supplement for details), before participants were debriefed about the background of the experiment. Participants took on average 290.61 seconds (*SD* = 156.00) to complete the questionnaire, with a mean time of 42.67 seconds (*SD* = 49.78) spent on the page that presented the experimental stimulus. We complied with APA (American Psychological Association) ethical standards in the treatment of our sample.

Stimuli

Participants read a short text about an innovative robot named Ellix. Based on our between-subject design, three versions of this vignette text were prepared. In the first condition, Ellix was introduced as a thought detector robot. In the second condition, Ellix was supposedly able to detect humans' emotions. In the third condition, the robot did not have any advanced analysis abilities, merely serving as a daily life tool. The descriptions were based on extracts of the mind perception classification by Gray et al. (2007); however, we made sure to highlight that the robot was not able to *feel/think* as was the focus of previous work (Appel et al., 2020; Gray & Wegner, 2012) but to *recognize* thinking or feeling on the human users' side. The stimuli texts were as follows (thought detector condition, emotion detector condition, control condition):

> *Ellix, a robot that can read your thoughts*
>
> Ellix is a social robot, i.e., a robot that is meant to interact with humans. Ellix is equipped with over 100 sensors and an advanced artificial intelligence system to make sense of the data it receives from its surroundings. It observes the human iris, facial expressions, voice patterns, and micro-movements of the head. It further studies the posture and movement of all other parts of the body. With decades worth of psychological insight stored in its algorithms, as well as machine learning procedures that make the system smarter with each use, Ellix is able to analyze human interaction partners. More specifically, Ellix possesses the constantly advancing ability to detect what humans think, for example which actions they wish to execute and whether or not they know the answer to a question.

> *Ellix, a robot that can read your emotions*
>
> Ellix is a social robot, i.e., a robot that is meant to interact with humans. Ellix is equipped with over 100 sensors and an advanced artificial intelligence system to make sense of the data it receives from its surroundings. It observes the human iris, facial expressions, voice patterns, and micro-movements of the head. It further studies the posture and movement of all other parts of the body. With decades worth of psychological insight stored in its algorithms, as well as machine learning procedures that make the system smarter with each use, Ellix is able to analyze human interaction

partners. More specifically, Ellix possesses the constantly advancing ability to detect what humans feel, for example which feelings they wish to act upon and whether or not they feel anxious when they answer a question.

Ellix, a robot with 100 sensors

Ellix is a social robot, i.e., a robot that is meant to interact with humans. Ellix is equipped with over 100 sensors and an advanced artificial intelligence system to make sense of the data it receives from its surroundings. It observes the human iris, facial expressions, voice patterns, and micro-movements of the head. It further studies the posture and movement of all other parts of the body. By these means, the system is equipped with the most recent technology to be useful as a daily-life tool.

Measures

Eeriness. The first dependent variable asked about users' feelings of eeriness in response to the robot and was measured with the help of three items ("uneasy," "unnerved," "creeped out") based on previous research (Gray & Wegner, 2012). A 7-point scale ranging from *not at all* (1) to *extremely* (7) was provided (Cronbach's α = .90, *M* = 3.61, *SD* = 1.83).

Concerns about human identity. This dependent variable was a composite of the repulsion scale (Kamide et al., 2012, two items) and three items of the concerns about human identity scale by Stein et al. (2019). These five items[2] were presented on a 7-point scale ranging from *strongly disagree* (1) to *strongly agree* (7), Cronbach's α = .91, *M* = 2.93 (*SD* = 1.59).

General evaluation. The third dependent variable consisted of three bipolar items ("hate it – love it;" "negative – positive;" "repulsive – attractive," Appel et al., 2019), which

[2] Items can be found in Appendix A.

were presented on a 7-point scale ranging from –3 to +3, Cronbach's α = .97, *M* = 0.43 (*SD* = 1.67).

Manipulation check. We asked participants to select the robot's ability that was introduced in the text describing the robot Ellix. Participants had to choose one of three options reflecting the description of the robot (see supplement for details).

Results

All *p*-values in this manuscript are based on two-tailed testing. Omnibus tests for the effects of the experimental manipulation on the three outcome variables were conducted. Pillai's Trace showed that the general linear model combining all three dependent variables did not reach statistical significance, *V* = 0.03, *F*(6, 662) = 1.89, *p* = .081, η_p^2 = .02. On closer inspection, between-subject tests showed a significant group difference for the dependent variable eeriness, *F*(2, 332) = 3.60, *p* = .028, η_p^2 = .021. Concerns about human identity, *F*(2, 332) = 1.27, *p* = .282, η_p^2 = .008, and participants' general evaluation of the robots, *F*(2, 332) = 2.56, *p* = .079, η_p^2 = .015, on the other hand, appeared to be unaffected by the treatment (see Table 3).

Table 3

Descriptive Results of Experiment 1

Variable	Thought detector		Emotion detector		Tool robot	
	M	*SD*	*M*	*SD*	*M*	*SD*
Eeriness	4.01	1.95	3.37	1.68	3.50	1.81
Concerns about human identity	3.05	1.60	2.73	1.51	3.00	1.64
General evaluation	0.12	1.81	0.50	1.63	0.60	1.57

Note. Sample sizes: Thought detector: *n* = 101, Emotion detector: *n* = 105, Tool robot: *n* = 129.

To test our specific hypotheses, planned contrasts were performed. As expected in Hypothesis 1a, the thought detector robot evoked higher eeriness than the emotion detector robot, *t*(332) = –2.53, *p* = .012, *d* = 0.35. The eeriness scores in response to the robot without analysis abilities (tool robot) were lower than the eeriness scores in the response to the thought detector, *t*(332) = 2.10, *p* = .036, *d* = 0.28, but they did not differ significantly from the emotion detector robot, *t*(332) = –0.56, *p* = .576, *d* = 0.07. Thus, the findings provide mixed support for Hypothesis 1b. An analysis contrasting the thought detector with both other conditions, *t*(332) = –2.65, *p* = .008, *d* = 0.31, underscores this pattern of results, indicating that the thought detector robot was perceived to be particularly eerie whereas the difference between the emotion detector robot and the control condition remained negligible.

As indicated by the omnibus analysis of variance (ANOVA), concerns about human identity were not affected by the experimental manipulation. The largest difference between the groups—which emerged between thought detector and emotion detector robot—did not reach statistical significance, *t*(332) = –1.44, *p* = .150, *d* = 0.20. Thus, no support was found for Hypotheses 2a and 2b.

Similarly, we note that the general evaluation of the thought detector robot did not differ significantly from the emotion detector robot, *t*(332) = 1.66, *p* = .097, *d* = 0.23 (Hypothesis 3a). While the robot without analysis abilities was evaluated more positively than the thought detector robot, *t*(332) = –2.19, *p* = .030, *d* = 0.29, it did not differ significantly from the emotion detector robot, *t*(332) = –0.45, *p* = .657, *d* = 0.06. As such, our results offer mixed support for Hypothesis 3b. When contrasting the general evaluation of the thought detector with both other conditions, a significant effect emerged *t*(332) = 2.19, *p* = .029, *d* = 0.26.

Discussion

The results of this experiment show that a thought detector robot evokes less favorable responses than a robot that can detect human emotions or serves as a simple tool, particularly in terms of higher eeriness. Eeriness has been described as a reaction to something that seems unfamiliar, an entity that eludes the world we know and feel comfortable with (e.g., Jentsch, 1906/1997; Mori, 1970). As humans are not yet used to the notion of having their thoughts and plans read, this detection ability might indeed push a machine right into the uncanny valley. In contrast, an emotion-detecting robot was perceived to be as harmless as a simple tool in our study; participants felt mostly at ease with this hypothetical machine. In our interpretation, this may be explained by people's familiarity with the respective recognition processes—as well as participants' confidence that emotional displays can be regulated and coped with and, thus, remain fully under their control.

In a critical reflection on our study, we note that the manipulation check—despite being successful—indicated that several members of the control group had experienced difficulties identifying their condition. Furthermore, more than three dozen participants interchanged the description of the thought detector robot with the description of the experience detector robot. As a takeaway from these observations, we adapted the materials for our follow-up research by highlighting the important parts of the descriptions in a bold font (see supplement). Since the evaluation of the emotion detector robot had not differed significantly from the tool robot, we further omitted the tool condition in our second study. Moreover, we advanced the current project by focusing on interindividual differences as an important influence on users' reactions to mind-reading machines.

3.3 Experiment 2

The first aim of Experiment 2 was to replicate our main result of Experiment 1: We expected that a thought detector robot would again be perceived to be eerier than an emotion detector robot. Additionally, we decided to focus on the potential influence of dispositional factors regarding user responses to mind-reading robots. Previous work showed that stable individual differences can explain eeriness as a response to humanoid robots (e.g., Lischetzke et al., 2017; MacDorman & Entezari, 2015; Rosenthal-von der Pütten & Weiss, 2015). Therefore, we developed several hypotheses based on the HEXACO model of personality—one of the most often used models of basic personality structure (Moshagen et al., 2019), which consists of the factors honesty-humility, emotionality, extraversion, agreeableness, conscientiousness, and openness to experience.

Extraversion

Extraverted people feel positive about themselves, enjoy leading groups and social interactions, and they experience positive feelings of enthusiasm and energy (Lee & Ashton, 2009). Prior research showed that high extraversion predicted positive responses to robots (Esterwood & Robert, 2020; Mou et al., 2020; Santamaria & Nathan-Roberts, 2017). Given these results, we assumed that extraversion predicted more positive responses to detector robots as well. No differences between thought detector and emotion detector robots were formulated.

H4: Being extraverted is associated with weaker feelings of eeriness evoked by mind-detecting robots.

Openness to Experience

People who are open to experience take an interest in unusual ideas, become absorbed in the beauty of art and nature, and are interested in various domains of knowledge (Lee & Ashton, 2009). Openness was a predictor for the acceptance of new technologies in general (Korukonda, 2007; Nov & Ye, 2008), and some research showed that this trait predicted positive responses to robots (Conti et al., 2017; Morsünbül, 2019; Rossi et al., 2018, 2020; but see Müller & Richert, 2018). We therefore hypothesize that openness to experience predicts more positive responses to detection robots. No differences between thought detector and emotion detector robots were formulated.

H5: Being open to experience is associated with weaker feelings of eeriness evoked by mind-detecting robots.

Emotionality

Emotionality is described by the extent to which people experience fear of physical danger, experience anxiety in potentially stressful situations, need emotional support from others and feel empathy for others (Lee & Ashton, 2009). Some research in the context of social robotics has dealt with the conceptually related factor of neuroticism. Neuroticism correlated with a more negative attitude towards a robot (Müller & Richert, 2018). These findings suggest that emotionality would predict higher aversion against supposedly mind-reading robots. No differences between thought detector and emotion detector robots were formulated.

H6: Being emotional is associated with stronger feelings of eeriness evoked by mind-detecting robots.

Agreeableness

People scoring high on this dimension tend to forgive wrongs that they suffered, are able to control their temper and are willing to compromise and cooperate with others (Lee & Ashton, 2009). Agreeableness was a predictor of trust in an autonomous security robot (Lyons et al., 2020) and was associated with higher trust in machines in general (Chien et al., 2016). Moreover, a higher score on agreeableness correlated with keeping a lower interpersonal distance to robots (Takayama & Pantofaru, 2009). Based on these results, a negative relationship with eeriness was expected for both detector robots. No differences between thought detector and emotion detector robots were formulated.

H7: Being agreeable is associated with weaker feelings of eeriness evoked by mind-detecting robots.

Conscientiousness

Conscientious persons organize their surroundings, are disciplined, and strive for perfection in their tasks (Lee & Ashton, 2009). No correlation between conscientiousness and the attitude towards robots was found in previous research (Müller & Richert, 2018). However, more conscientious people rated robot motion more negatively than less conscientious persons (Bodala et al., 2020) and preferred a text interface compared to a virtual character (Looije et al., 2010). Given these few and mixed findings, we formulated no formal hypothesis and also no assumptions regarding differences between thought detector and emotion detector robots.

Honesty-Humility

The dimension Honesty-Humility is pronounced for people who avoid manipulating others for personal gain, who do not enjoy breaking rules and are uninterested in luxuries (Lee & Ashton, 2009). Special focus was put on the moderating role of the trait honesty-humility in our study. We assumed that people scoring high in the honesty-humility dimension would be less opposed to thought detection, as their overt behavior tends to be in line with their thoughts and plans. The latter is shown by negative correlations between honesty-humility and cheating behavior (Hilbig & Zettler, 2015; Kleinlogel et al., 2018, Moshagen et al., 2018; Pfattheicher et al., 2019). In human-robot interaction, cheating was negatively correlated with honesty-humility when a robot gave instructions (Petisca et al., 2019). Based on this line of argumentation, an interaction hypothesis was put forward.

> **H8:** Scoring low in the honesty-humility dimension increases the difference of eeriness evoked by the thought detector robot and the emotion detector robot.

Method

Participants

An a-priori power analysis with G*Power and considerations regarding power of moderation effects (Giner-Sorolla, 2018; Simonsohn, 2014) yielded an aspired sample size of 500 participants. We invited 600 people of the MTurk participant pool (US residence, hit approval rate > 98%, hits > 1000) to participate in our online experiment to have a buffer if careless responding occurred. Of the 602 completions, 20 participants did not have sufficient English skills and were therefore not included in the analyses (Kennedy et al., 2020). Five additional participants failed at least one attention check item and another eight participants had large (> ± 3 years) deviations when asked twice about their age. Moreover, 16 participants were excluded because their participation time was lower than 200 seconds (*n* = 10) or higher than 2800 seconds (*n* = 6). Seventeen participants interchanged the thought detector robot and the emotion detector robot, failing the manipulation check. The remaining sample consisted of 536 participants (238 female, 291 male, 7 non-binary or no answer) with an average age of 40.35 years (*SD* = 11.96, ranging from 19 to 79 years). Exploratory analyses revealed that age and gender did not moderate the influence of the robot manipulation on eeriness (see supplement).

Procedure

Again, we asked participants to give informed consent before starting the online experiment. Questions that allow conclusions to be drawn about data quality were included in a similar manner than in the first experiment (see supplement). Participants were randomly assigned to read a text about one of two robots: a thought detector robot or an emotion detector robot. The same stimuli as in Experiment 1 were used, albeit with a slight variation, we highlighted the manipulated parts of the descriptions in bold font (see supplement). As an improved manipulation check, participants had to select the abilities of the robot about which they had been informed immediately after reading the robot descriptions. Subsequently, the participants filled in the eeriness and HEXACO measures, followed by the negative attitude towards a robot scale (Nomura et al., 2006) which was used in an exploratory analysis (see supplement). The survey ended with sociodemographic questions, an opportunity to leave comments, and a debriefing. It took participants an average of 662.09 seconds (*SD* = 1063.97) to complete the questionnaire, including a mean duration of 65.24 seconds (*SD* =

106.44) spent on the page that presented the experimental stimulus. Again, we complied with APA ethical standards in the treatment of our sample.

Measures

Eeriness. Eeriness was measured with the three items used in Experiment 1, resulting in a mean of *M* = 3.72 (*SD* = 1.93), Cronbach's α = .91.

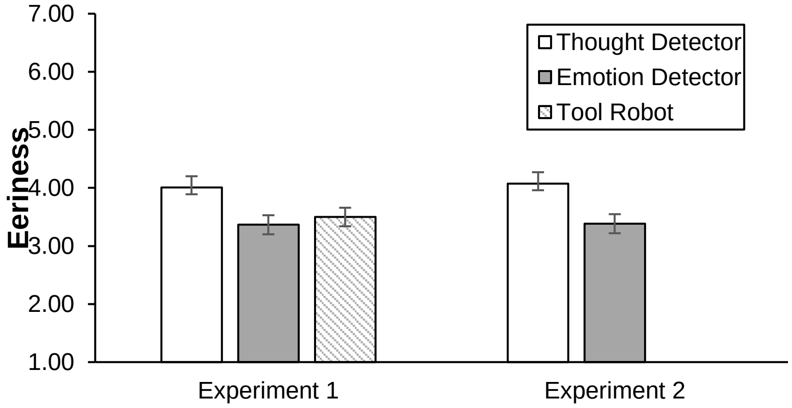
Personality. We used the HEXACO-60 questionnaire (Ashton & Lee, 2009), consisting of 60 items. Each dimension was measured through ten items on a 5-point scale ranging from 1 (*strongly disagree*) to 5 (*strongly agree*). All Cronbach's αs reached values of .72 or above. For detailed descriptive statistics see Table S1 and Table S2 in the supplement.

Results

In support of Hypothesis 1a and replicating the results of Experiment 1, the thought detector robot (*M* = 4.08, *SD* = 1.87) was perceived to be significantly eerier than the emotion detector robot (*M* = 3.38, *SD* = 1.92), *t*(534) = 4.23, *p* < .001, *d* = 0.37 (see Figure 1 eeriness results in both experiments).

Figure 1

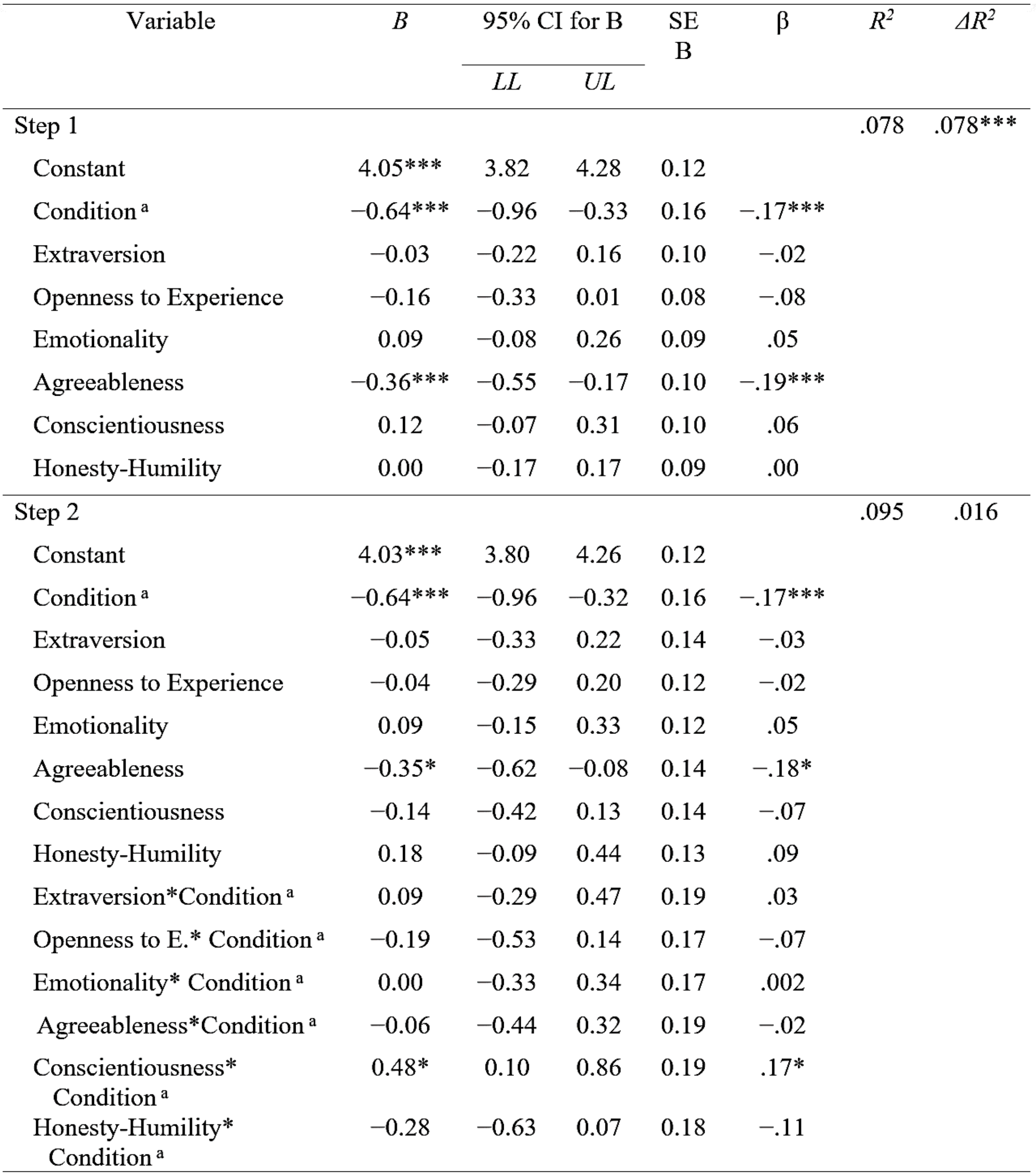
Eeriness Means and Standard Errors in Both Experiments



The main effects and interactions of robot condition and HEXACO dimensions were analyzed by a hierarchical two-step regression. The results of the regression model are depicted in Table 4.

Table 4

Results of a Hierarchical Regression Analysis



Variable	*B*	95% CI for B		SE B	β	*R*^{2}	*ΔR*^{2}
		LL	*UL*				
Step 1						.078	.078***
Constant	4.05***	3.82	4.28	0.12			
Condition [a]	−0.64***	−0.96	−0.33	0.16	−.17***		
Extraversion	−0.03	−0.22	0.16	0.10	−.02		
Openness to Experience	−0.16	−0.33	0.01	0.08	−.08		
Emotionality	0.09	−0.08	0.26	0.09	.05		
Agreeableness	−0.36***	−0.55	−0.17	0.10	−.19***		
Conscientiousness	0.12	−0.07	0.31	0.10	.06		
Honesty-Humility	0.00	−0.17	0.17	0.09	.00		
Step 2						.095	.016
Constant	4.03***	3.80	4.26	0.12			
Condition [a]	−0.64***	−0.96	−0.32	0.16	−.17***		
Extraversion	−0.05	−0.33	0.22	0.14	−.03		
Openness to Experience	−0.04	−0.29	0.20	0.12	−.02		
Emotionality	0.09	−0.15	0.33	0.12	.05		
Agreeableness	−0.35*	−0.62	−0.08	0.14	−.18*		
Conscientiousness	−0.14	−0.42	0.13	0.14	−.07		
Honesty-Humility	0.18	−0.09	0.44	0.13	.09		
Extraversion*Condition [a]	0.09	−0.29	0.47	0.19	.03		
Openness to E.* Condition [a]	−0.19	−0.53	0.14	0.17	−.07		
Emotionality* Condition [a]	0.00	−0.33	0.34	0.17	.002		
Agreeableness*Condition [a]	−0.06	−0.44	0.32	0.19	−.02		
Conscientiousness* Condition [a]	0.48*	0.10	0.86	0.19	.17*		
Honesty-Humility* Condition [a]	−0.28	−0.63	0.07	0.18	−.11		

Note. All continuous predictors were z-standardized; *N* = 536; CI = Confidence Interval; *LL* = lower limit; *UL* = upper limit.

[a] dummy-coded (0 – thought detector robot; 1 – emotion detector robot).

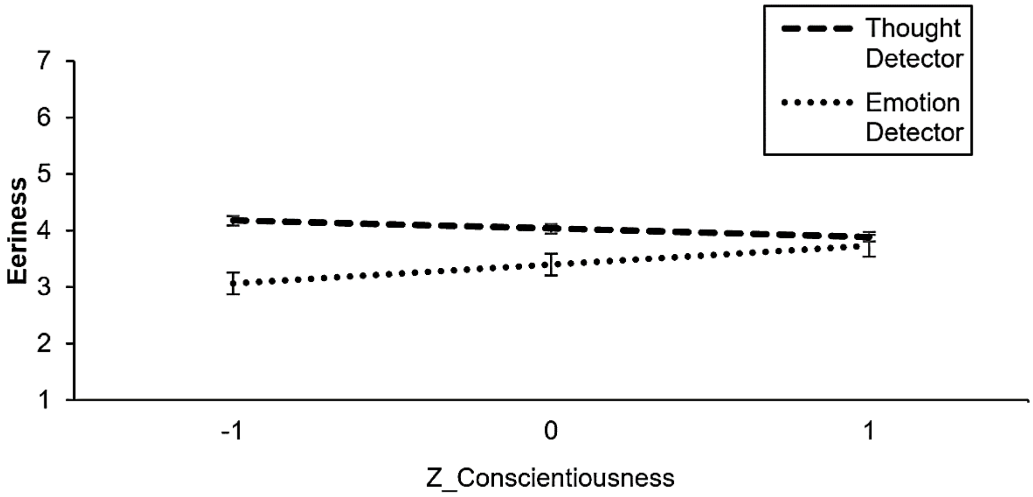
p* < .05. * *p* < .001.

In the first step of the hierarchical regression, all six HEXACO traits and the experimental factor were entered. In addition to the main effect of the experimental factor, a significant effect was found for agreeableness, $t(530) = -3.74$, $p < .001$. As expected in Hypothesis 7, being agreeable was associated with a lower level of eeriness evoked by the detector robots. None of the assumed remaining HEXACO effects reached statistical significance, so Hypotheses 4, 5, and 6 had to be rejected.

The second regression step—which also included interaction terms between the HEXACO dimensions and the assigned condition—revealed no interaction effect for honesty-humility, which led to a rejection of Hypothesis 8. However, unexpectedly, we observed a significant interaction between participants' conscientiousness and the robot condition, $B = .48$, $SE = 0.19$, $p = .014$, $\Delta R^2 = .01$ (see Figure 2), which was further examined using the SPSS-macro PROCESS (Hayes, 2012). Follow-up analyzes (Aiken & West, 1991) revealed that participants who were low in conscientiousness (−1 *SD*) perceived the thought detector robot to be significantly eerier than the emotion detector, $B = -1.11$, $SE = 0.25$, $t(524) = -4.45$, $p < .001$, 95% CI [−1.61, −0.62]. In contrast, the detector condition had no impact on participants who were high in conscientiousness (+1 *SD*), $B = -0.16$, $SE = 0.25$, $t(524) = -0.64$, $p = .519$, 95% CI [−0.66, 0.33]. According to the Johnson-Neyman technique, the manipulation of detecting abilities had a significant effect on participants' perceived eeriness for z-standardized values ≤ 0.54 of conscientiousness. About 69.59% of our participants fell into this significant region.

Figure 2

Interaction between Robot and Conscientiousness



Note. Error bars represent ± 1SE.

Discussion

Corroborating our results from Experiment 1, the thought detector robot was perceived as significantly eerier than the emotion detector robot. Moreover, a significant effect of agreeableness was found: Higher levels in this basic personality dimension were associated with less eeriness ascribed to the detector robots, matching the way this trait had affected user responses in prior human–robot studies (e.g., Chien et al., 2016; Lyons et al., 2020; Takayama & Pantofaru, 2009). As people high in agreeableness typically react in a tolerant and kind-mannered way to outside influences, it comes as little surprise that they also responded more positively to the presented detection robots. At the same time, we were surprised by a lack of noteworthy effects for the remaining HEXACO dimensions. Also, unlike expected, our data did not reveal a significant interaction of the dimension honesty-humility and the robot condition in our moderated regression analysis. Instead, the thought detector robot was generally evaluated as eerier than the emotion detector robot, regardless of participants' honesty-humility scores. As a main result of our second experiment, we therefore note that people’s evaluation of detector robots appears to be mostly unaffected by their fundamental personality traits. Arguably, this implies that the notion of sophisticated analysis robots may cause unease in a rather universal way, emerging as a strong challenge to people’s idea of a good, unthreatening machine.

It should be noted, however, that our data yielded an unexpected interaction effect regarding another HEXACO trait: The higher participants scored in conscientiousness, the smaller was the difference between the eeriness ratings for the two detector robots. In our interpretation, this might be explained by the specific characteristics of highly conscientious individuals, who tend to put a strong emphasis on (cognitive) achievement and performance, while considering overt emotions as detrimental for success (Witteman et al., 2009; for an overview of the interplay of conscientiousness and negative affect see Fayard et al., 2012; Javaras et al., 2012). Further research is needed to find out how human conscientiousness influences interactions with robots—and to scrutinize the robustness of the uncovered interaction effect.

3.4 General Discussion

Robots and artificial intelligence are considered key technologies for the societies of today–even if not all prophecies made in science fiction have materialized (yet). User responses to these advanced technologies are of basic and applied relevance. Connecting the mind perception literature (Gray et al., 2007) and the uncanny valley hypothesis (Jentsch, 1906/1997; Mori, 1970), research on human-machine-interactions has demonstrated that robots who are ascribed human mind elicit negative responses such as eeriness (e.g., Stein & Ohler, 2017). Importantly, machines with emotions (experience) were found to be more aversive (Appel et al., 2020; Gray & Wegner, 2012; Taylor et al., 2020) than machines with thoughts and plans (agency). Unlike previous research that was primarily focused on user responses to mind in a machine, we focused on a reversed perspective—the evaluation of machines capable of reading the human mind. Following our data analysis, we report that our main assumption held true across two experiments: In the realm of *mind-reading* machines, a thought detector is perceived as eerier than an emotion detector. With this fascinating outcome, we suggest that our results clearly advance the investigation of the *uncanny valley of mind* (Kang & Sundar, 2019; Stein & Ohler, 2017), both by shifting its overarching perspective and by introducing an important cognitive component. Offering further support for this main result, our second experiment showed that the stronger aversion against thought-detecting machines remained independent of several basic HEXACO personality dimensions. To us, this suggests that being apprehensive towards the concept of thought detection connects most humans regardless of their personality dispositions.

Proceeding to a psychological interpretation of our findings, we suggest that the need to perceive oneself as being in control is as important for human-robot interactions as it is for human-human interactions; potentially even more so. This desire for control, however, may be harmed by robots that appear able to look into the human mind. While we are used to sharing (and hiding) our emotions during many daily life interactions, it turns into a much more delicate matter if robots or other AI-based systems start to correctly infer what its user is thinking; in a dystopian scenario, this information could quickly be used against the human user in question, for instance in a job assessment or law-related context. Considering that the fear of artificial intelligence turning against humans has been named as a central caveat of human-computer interaction research (Cave & Dihal, 2019), even the most pessimistic imaginations should probably be kept in mind when designing detector robots. Based on our findings, we recommend that developers of robotic and AI systems strive for absolute transparency regarding the capabilities of their created products and machines. Privacy guidelines should always be incorporated to make sure that the detecting entity does not share the results of its analysis with third parties; in all likability, this will help to alleviate the apprehension among potential users.

Limitations and Future Work

We note several limitations of the current experiments, which might also offer inspiration for future work. First, the observed mean eeriness ratings ranged between 3 and 4 on a 7-point scale, implying that the robot descriptions did not elicit particularly strong eeriness among participants. We assume that the online survey methodology paired with written text manipulations increased participants' psychological distance to our stimuli, thus preventing stronger emotional reactions. Similarly, since we (purposely) did not offer any information about the robots' appearance, some participants might have imagined a very friendly-looking or cute machine, which might have "softened" the eeriness evoked by our mind manipulation.

Second, we did not specify which emotions or thoughts could be detected by the robot. Emphasizing the detection of *negative* feelings or cognitions, for example, could have increased eeriness ratings in a notable manner, as participants might see it as more discomforting to have their sadness, anxiety, or anger discovered. A similar notion concerns the reading of thoughts, as it appears highly likely that some cognitions might be more sensitive or confidential for us than others. Hence, future research is encouraged to examine differences in users' experience and evaluations in response to robots detecting different thoughts and emotions.

Lastly, we believe that the methodological approach of using written vignette texts as stimuli deserves particular attention. While we still consider it as a very useful way of putting the mental abilities of a machine front and center, it might be worth considering to also show pictures or even focus on live interactions with real robots in order to advance the discussed line of research. Doing so, fascinating interaction effects between the robots' mental capacities and its specific embodiment could be found, as suggested by another recent study (Stein et al., 2020). Building upon this, the influence of thought detection or emotion detection could also be explored in very different contexts: For instance, we strongly believe that a robot's capability to detect aspects of human mind will be evaluated differently in court cases, a therapeutic setting, nursing scenarios or smart homes (Thakur & Han, 2018). This way, evidence on the generalizability of the reported main effect could be gathered. Along the same lines, it should be explored whether the stronger aversion against a thought-detecting machine also persists in other cultures, as all participants taking part in our online experiments were recruited in the United States. Specifically, it might make sense to focus on participants from more collectivistic societies in future efforts, as the stronger social interdependence in the respective countries might also modulate the desire to avoid having one's mind read by another entity.

Conclusion

As cherished in the German folk song mentioned at the beginning of this paper (*Die Gedanken sind frei*), humans seem to truly appreciate the fact that their thoughts may roam free, without the risk of insulting others or having to admit one's secret desires. Accordingly, we found that the concept of thought-detecting machines—a hypothetical notion that does not seem so far removed from reality anymore, considering current technical developments—elicits significantly more unease than the concurrent idea of a robot analyzing human feelings. While this psychological observation may give developers pause or make them question the ethical boundaries of their innovations, it may also be possible to pave a path for well-accepted thought detectors; as long as control perceptions are kept in mind, people might get used to this novel experience after all.

3.5 Supplementary Material

“Mind-Reading Machines: Distinct User Responses to Thought-Detecting and Emotion-Detecting Robots”

Supplement 1 — Experiment 1: Sequence and Wording of Hypotheses

Supplement 2: Robot Descriptions

Supplement 3 — Experiment 1: Manipulation Check

Supplement 4 — Experiment 1: Questions and Actions to Ensure Data Quality

Supplement 5 — Experiment 2: Questions and Actions to Ensure Data Quality

Supplement 6 — Experiment 2: Descriptives of the HEXACO Scales

Supplement 7 — Experiment 2: Correlations of the Predictors

Supplement 8 — Experiment 2: Statistical Analysis Including the Covariate NARS

Supplement 9: Exploratory Analyses Considering Age and Gender

Supplement References

Supplement 1 — Experiment 1: Sequence and Wording of Hypotheses

Pre-registration	Manuscript
H1a: The thought detector robot will evoke higher eeriness than the emotion detector robot. H2a: The robot without analysis abilities (control group) will evoke the least eeriness.	H1: The thought detector robot will evoke higher eeriness than the emotion detector robot (H1a), whereas the robot without analysis abilities will evoke the least eeriness (H1b).
H1b: The thought detector robot will evoke stronger concerns about human identity than the emotion detector robot. H2b: The robot without analysis abilities (control group) will evoke the weakest concerns about human identity.	H2: The thought detector robot will evoke stronger concerns about human identity than the emotion detector robot (H2a), whereas the robot without analysis abilities will evoke the least concerns (H2b).
H1c: The thought detector robot will be evaluated as a more negative new technology than the emotion detector robot. H2c: The robot without analysis abilities (control group) will be evaluated as the most positive new technology.	H3: The thought detector robot will yield a lower general evaluation of the new technology than the emotion detector robot (H3a), whereas the robot without analysis abilities will yield the best evaluation of the new technology (H3b).

Supplement 2 — Experiment 2: Robot Descriptions

Thought Detector (emphases made in bold only in Experiment 2)

Ellix, a robot that can read your thoughts

Ellix is a social robot, i.e., a robot that is meant to interact with humans. Ellix is equipped with over 100 sensors and an advanced artificial intelligence system to make sense of the data it receives from its surroundings. It observes the human iris, facial expressions, voice patterns, and micro-movements of the head. It further studies the posture and movement of all other parts of the body. With decades worth of psychological insight stored in its algorithms, as well as **machine learning procedures** that make the system smarter with each use, Ellix is able to **analyze human interaction partners**. More specifically, Ellix possesses the constantly advancing ability to **detect what humans think**, for example **which actions they wish to execute** and whether or not they know the answer to a question.

Emotion Detector (emphases made in bold only in Experiment 2)

Ellix, a robot that can read your emotions

Ellix is a social robot, i.e., a robot that is meant to interact with humans. Ellix is equipped with over 100 sensors and an advanced artificial intelligence system to make sense of the data it receives from its surroundings. It observes the human iris, facial expressions, voice patterns, and micro-movements of the head. It further studies the posture and movement of all other parts of the body. With decades worth of psychological insight stored in its algorithms, as well as **machine learning procedures** that make the system smarter with each use, Ellix is able to **analyze human interaction partners**. More specifically, Ellix possesses the constantly advancing ability to **detect what humans feel**, for example **which feelings they wish to act upon and whether or not they feel anxious when they answer a question**.

Tool Robot (= Control Group in Experiment 1)

Ellix, a robot with 100 sensors

Ellix is a social robot, i.e., a robot that is meant to interact with humans. Ellix is equipped with over 100 sensors and an advanced artificial intelligence system to make sense of the data it receives from its surroundings. It observes the human iris, facial expressions, voice patterns, and micro-movements of the head. It further studies the posture and movement of all other parts of the body. By these means, the system is equipped with the most recent technology to be useful as a daily-life tool.

Supplement 3 — Experiment 1: Manipulation Check

As a manipulation check, we asked participants to select the robot's ability that was introduced in the text describing the robot Ellix. Three options were provided that represented the three conditions, 1) "The robot has the capacity to detect what humans think and which actions they wish to execute," 2) "The robot has the capacity to detect what humans feel and whether or not they feel anxious answering questions," 3) "The robot does not have the capacity to analyze humans' thoughts and feelings." Participants had to choose one of the options.

We excluded 39 participants who read the thought detector vignette and answered that they had read a text describing the emotion detector robot (Response Option 2) or who read the emotion detector vignette and answered that they had read a text describing the thought detector robot (Response Option 1). This criterion was not pre-registered.

A Chi-Square-Test showed that the participants of our final sample were able to correctly identify the robot condition they had been assigned to when a short extract of their treatment text was shown again, $\chi^2(4, N = 335) = 256.27$, $p < .001$, $\varphi = .88$.

Supplement 4 — Experiment 1: Questions and Actions to Ensure Data Quality

- Study description in correct English: *Please describe the study you are currently participating in. Please use full English sentences. Do not use more than 50 words.*
- Language check: Picture of the fruit „eggplant“, *How do you call this food?*
- Captcha: *Please confirm you are not a bot*
- Attention Check: *What was this study about?* (innovative robots, climate change, fake news)?
- Attention Check: *“Please click strongly agree here”*
- Processing time (min. 100s, max. 920s)
- Comparison of entered age with given year of birth

Supplement 5 — Experiment 2: Questions and Actions to Ensure Data Quality

- Study description in correct English: *Please describe the study you are currently participating in. Please use full English sentences. Do not use more than 50 words.*
- Attention Check: *What was this study about?* (innovative robots, climate change, fake news)?
- Attention Check: *“Please click strongly agree here”*
- Processing time (min. 200s, max. 2800s)
- Comparison of entered age with given year of birth

Supplement 6 — Experiment 2: Descriptives of the HEXACO Scales

Table S1

Scale Statistics of the HEXACO Questionnaire

Dimension	Example item	*M*	*SD*	Cronbach's α
Honesty-Humility	"I wouldn't use flattery to get a raise or promotion at work, even if I thought it would succeed."	3.59	0.78	.82
Emotionality	"I sometimes can't help worrying about little things."	3.18	0.78	.84
Extraversion	"I feel reasonably satisfied with myself overall."	3.13	0.84	.87
Agreeableness	"I rarely hold a grudge, even against people who have badly wronged me."	3.33	0.77	.85
Conscientiousness	"I plan ahead and organize things, to avoid scrambling at the last minute."	3.84	0.58	.72
Openness to Experience	"I would enjoy creating a work of art, such as a novel, a song, or a painting."	3.67	0.75	.82

Supplement 7 — Experiment 2: Correlations of the Predictors

Table S2

Zero-order Correlations of the Predictors

Variable	1	2	3	4	5	6	7
1.Condition[a]	–	.07	.00	−.08	.07	.07	.00
2.Extraversion		–	.17***	−.29***	.41***	.40***	.06
3.Openness to Experience			–	.00	.18***	.28***	.07
4.Emotionality				–	−.19***	−.12**	−.15***
5.Agreeableness					–	.41***	.32***
6.Conscientiousness						–	.27***
7.Honesty-Humility							–

Note. Sample size: *N* = 536.

[a] dummy-coded (0 = thought detector robot; 1 = emotion detector robot).

** *p* <.01. *** *p* < .001.

Supplement 8 — Experiment 2: Statistical Analyses Including the Covariate NARS

The three subscales developed by Nomura et al. (2006) were presented after the HEXACO scales in the questionnaire and averaged to create an overall negative attitude toward robots scale (NARS) score (*M* = 2.71, *SD* = 0.80). The 14 items (e.g., “I would feel uneasy if robots really had emotions”) were presented on a 5-point scale ranging from 1 (*strongly disagree*) to 5 (*strongly agree*), Cronbach’s α = .91. Zero-order correlations including the NARS as a covariate can be found in Table S3, while Table S4 shows the results of the regression approach.

Table S3

Zero-order Correlations of the Predictors Including the Covariate NARS

Variable	1	2	3	4	5	6	7	8
1.Condition[a]	−	.07	.00	−.08	.07	.07	.00	−.12**
2.Extraversion		−	.17***	−.29***	.41***	.40***	.06	−.18***
3.Openness to Experience			−	.00	.18***	.28***	.07	−.23***
4.Emotionality				−	−.19***	−.12***	−.15***	.14**
5.Agreeableness					−	.41***	.32***	−.32***
6.Conscientiousness						−	.27***	−.19***
7.Honesty-Humility							−	−.12**
8.NARS								−

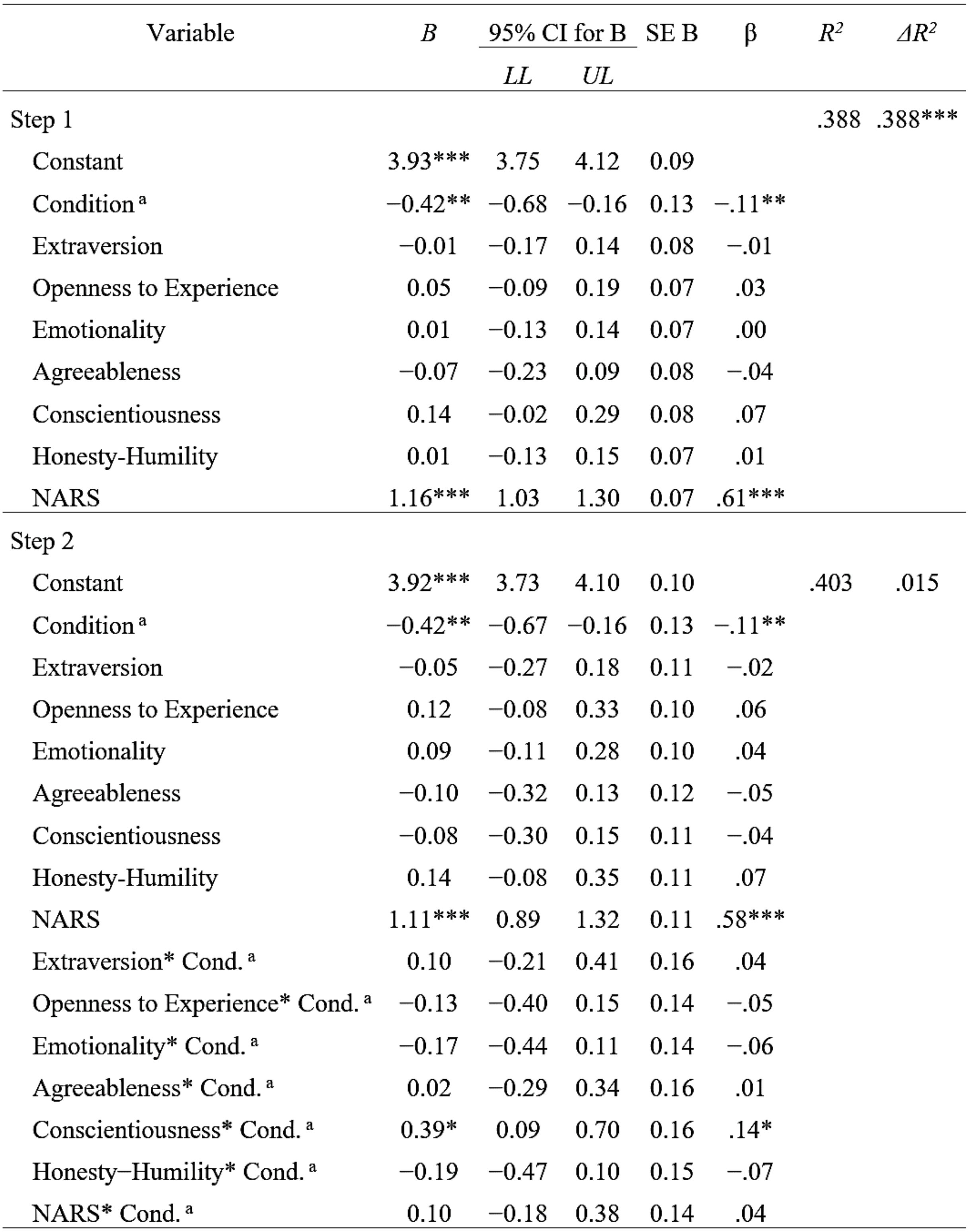
Note. Sample size: *N* = 536.

[a] dummy-coded (0 – thought detector robot; 1 – emotion detector robot).

** p <.01. *** *p* < .001.

Table S4

Results of the Hierarchical Regression Including the Covariate NARS



Variable	*B*	95% CI for B		SE B	β	*R*²	*ΔR*²
		LL	*UL*				
Step 1						.388	.388***
Constant	3.93***	3.75	4.12	0.09			
Condition [a]	−0.42**	−0.68	−0.16	0.13	−.11**		
Extraversion	−0.01	−0.17	0.14	0.08	−.01		
Openness to Experience	0.05	−0.09	0.19	0.07	.03		
Emotionality	0.01	−0.13	0.14	0.07	.00		
Agreeableness	−0.07	−0.23	0.09	0.08	−.04		
Conscientiousness	0.14	−0.02	0.29	0.08	.07		
Honesty-Humility	0.01	−0.13	0.15	0.07	.01		
NARS	1.16***	1.03	1.30	0.07	.61***		
Step 2							
Constant	3.92***	3.73	4.10	0.10		.403	.015
Condition [a]	−0.42**	−0.67	−0.16	0.13	−.11**		
Extraversion	−0.05	−0.27	0.18	0.11	−.02		
Openness to Experience	0.12	−0.08	0.33	0.10	.06		
Emotionality	0.09	−0.11	0.28	0.10	.04		
Agreeableness	−0.10	−0.32	0.13	0.12	−.05		
Conscientiousness	−0.08	−0.30	0.15	0.11	−.04		
Honesty-Humility	0.14	−0.08	0.35	0.11	.07		
NARS	1.11***	0.89	1.32	0.11	.58***		
Extraversion* Cond. [a]	0.10	−0.21	0.41	0.16	.04		
Openness to Experience* Cond. [a]	−0.13	−0.40	0.15	0.14	−.05		
Emotionality* Cond. [a]	−0.17	−0.44	0.11	0.14	−.06		
Agreeableness* Cond. [a]	0.02	−0.29	0.34	0.16	.01		
Conscientiousness* Cond. [a]	0.39*	0.09	0.70	0.16	.14*		
Honesty–Humility* Cond. [a]	−0.19	−0.47	0.10	0.15	−.07		
NARS* Cond. [a]	0.10	−0.18	0.38	0.14	.04		

Note. All continuous predictors were z-standardized; *N* = 536; CI = Confidence Interval; *LL* = lower limit; *UL* = upper limit.

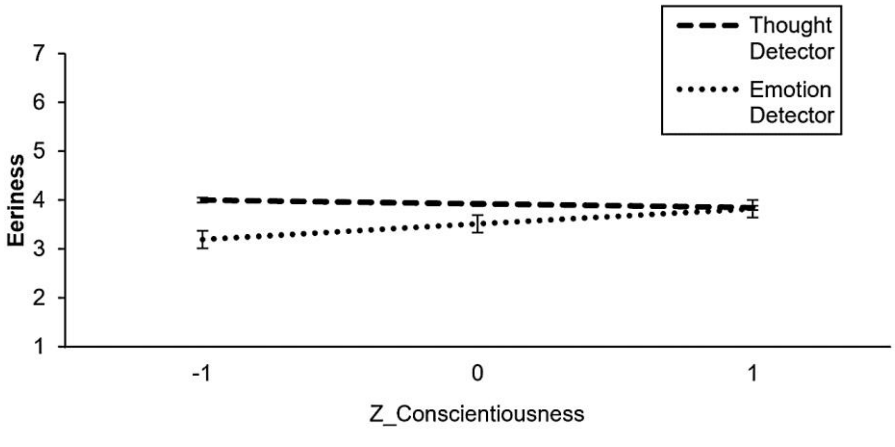
[a] dummy-coded (0 – thought detector robot; 1 – emotion detector robot).

p* < .05. ***p* < .01. * *p* < .001.

When adding the NARS as a covariate in the hierarchical regression model, no significant effect was found for agreeableness (as in the original analysis), but for the NARS, *t*(529) = 10.21, *p* < .001. Also, the robot condition remained a significant predictor. The significant interaction effect (see Figure S1) between conscientiousness and robot condition persisted, *B* = .39, *SE* = 0.16, *p* = .013, ΔR^2 = .01. Follow-up analyses (Aiken & West, 1991) revealed that participants who were low in conscientiousness (−1 *SD*) perceived the thought detector robot to be significantly eerier than the emotion detector, *B* = −0.81, *SE* = 0.20, *t*(522) = −3.94, *p* < .001, 95% CI [−1.21, −0.40]. In contrast, the detector condition had no impact on participants who were high in conscientiousness (+1 *SD*), *B* = −0.03, *SE* = 0.21, *t*(522) = −0.13, *p* = .894, 95% CI [−0.43, 0.38]. According to the Johnson-Neyman technique, the manipulation of detecting abilities had a significant effect on participants' perceived eeriness for z-standardized values ≤ 0.35 of conscientiousness. About 61.94% of our participants fell into this significant region.

Figure S1

Interaction of Robot Group and Conscientiousness Including the Covariate NARS



Note. Error bars represent ± 1SE.

Supplement 9: Exploratory Analyses Considering Gender and Age

We collected age and gender to describe the demographic characteristics of our sample. We performed some exploratory analyses which are reported below. In the following analyses, participants with non-binary gender were excluded from the analyses.

In Experiment 1, gender was equally distributed across conditions, $\chi^2(2, N = 330) = 0.29$, $p = .866$, φ = .03. Adding gender as an additional factor in the multivariate analysis of variance (MANOVA), we find a significant main effect of gender on concerns about human identity, $F(1, 324) = 5.38$, $p = .021$, $\eta_p^2 = .02$, but no interaction effect with the robot group, $F(1, 324) = 0.83$, $p = .437$, $\eta_p^2 = .01$. Additionally, gender correlated positively with concerns about human identity ($r = .13$, $p = .020$), indicating that female participants slightly reported more concerns about human identity. We found no significant correlations of gender with eeriness ($r = .05$, $p = .403$) or the evaluation of the technology ($r = –.09$, $p = .107$).

The distribution of age slightly differed across conditions in Experiment 1, $F(2, 327) = 5.29$, $p = .005$, $\eta_p^2 = .03$. Age was neither a predictor of eeriness, $B = −0.01$, $SE = 0.02$, $t(328) = −0.77$, $p = .444$, nor the concerns about human identity, $B = −0.01$, $SE = 0.01$, $t(328) = −0.61$, $p = .542$, or the evaluation of the technology, $B = −0.01$, $SE = 0.01$, $t(328) = −0.46$, $p = .645$. No significant interactions of robot group and age were found.

We re-ran the main analyses with age and gender as control variables. The results remained virtually the same. More specifically, we found the reported difference between the thought detector robot and the emotion detector robot on eeriness, $F(1, 198) = 5.69$, $p = .018$, $\eta_p^2 = .03$. The eeriness scores in response to the robot without analysis abilities (tool robot) were marginally lower than the eeriness scores in the response to the thought detector, $F(1, 224) = 3.83$, $p = .052$, $\eta_p^2 = .02$, but they did not differ significantly from the emotion detector robot, $F(1, 226) = 0.29$, $p = .589$, $\eta_p^2 = .001$. The largest difference in the dependent variable concerns about human identity was found between the thought detector robot and the emotion detector robot but did not reach statistical difference, $F(1, 198) = 3.57$, $p = .060$, $\eta_p^2 = .02$. Again, comparable to the original analysis, the tool robot was evaluated as a more positive new technology than the thought detector robot, $F(1, 224) = 4.18$, $p = .042$, $\eta_p^2 = .02$.

In Experiment 2, gender was also equally distributed across conditions, $\chi^2(1, N = 529) = 2.95$, $p = .086$, $\varphi = –.08$. A 2×2 ANOVA with gender and robot condition as independent variables and eeriness as a dependent variable revealed a significant effect of the robot condition, $F(1, 525) = 16.45$, $p < .001$, $\eta_p^2 = .03$, but neither a significant main effect of gender, $F(1, 525) = 2.87$, $p = .091$, $\eta_p^2 = .01$, nor a significant interaction effect, $F(1, 525) = 2.68$, $p = .102$, $\eta_p^2 = .01$.

Age was equally distributed across conditions in Experiment 2, $t(525) = –0.40$, $p = .691$. Age did not predict eeriness when added as a z-standardized-predictor in the hierarchical regression approach in addition to the robot condition and HEXACO dimensions, $B = 0.08$, $SE = 0.09$, $t(516) = 0.91$, $p = .362$. No interaction of age and robot condition was found in Experiment 2 using the PROCESS-Makro, $B = 0.03$, $SE = 0.18$, $t(515) = 0.18$, $p = .860$.We further re-ran the analyses with age and gender as control variables. The results remained virtually unchanged. The thought detector robot was perceived to be significantly eerier than the emotion detector robot, $F(1, 523) = 18.02$, $p < .001$, $\eta_p^2 = .03$.

Supplement References

Aiken, L. S., & West, S. G. (1991). *Multiple regression: Testing and interpreting interactions.* Sage.

Nomura, T., Kanda, T., & Suzuki, T. (2006). Experimental investigation into influence of negative attitudes toward robots on human–robot interaction. *AI & Society, 20*(2), 138-150. https://doi.org/10.1007/s00146-005-0012-7

4. Project 2 | Improving Evaluations of Advanced Robots By Depicting Them in Harmful Situations

Andrea Grundke

Jan-Philipp Stein

Markus Appel

Status:

Published in the journal *Computers in Human Behavior*

Formal Citation/Reference:

Grundke, A., Stein, J.-P., & Appel, M. (2023). Improving evaluations of advanced robots by depicting them in harmful situations. *Computers in Human Behavior, 140,* Article 107565. https://doi.org/10.1016/j.chb.2022.107565

Highlights

- Robots with mind evoked aversion in prior studies (uncanny valley of mind)
- Two online experiments were conducted, testing human empathy as a countermeasure
- A robot shown in a harmful situation elicited higher empathy
- An indirect effect of the situation on likeability, via empathy, was observed
- A negative residual effect of showing the robot in a harmful situation emerged

Abstract

Equipping robots with sophisticated mental abilities can result in reduced likeability (*uncanny valley of mind*). Other work shows that exposing robots to harm increases empathy and likeability. Connecting both lines of research, we hypothesized that eliciting empathy could mitigate or even reverse the negative response to robots with mind. In two online experiments, we manipulated the attributes of a robot (with or without mind) and presented the robot in situations in which it was either exposed to harm or not. Perceived empathy for the robot and robot likeability served as dependent variables. Experiment 1 (*N* = 559) used text vignettes to manipulate robot mind and a video that involved either physical harm or no harm to the machine. In a second experiment (*N* = 396), both experimental factors were manipulated via the shown video. Across both experiments, we observed a significant indirect effect of presenting the robot in a harmful situation on likeability, with empathy serving as a mediating variable. Moreover, a residual negative influence of showing the robot in a harmful situation was detected. We conclude that the uncanny valley of mind observed in our studies could be based on the robot's human-like imperfection, rather than descriptions of its supposed mind.

Keywords: uncanny valley, mind perception, empathy, human-robot interaction, user acceptance

4.1 Introduction

Due to rapid technological advancements, robotic technology has become much more complex, diverse, and visible in recent years (International Federation of Robotics, 2021; Yang et al., 2020). At the same time, people's attitudes towards robots are not only ambivalent (Brondi et al., 2021; Stapels & Eyssel, 2022), empirical data suggest that attitudes towards robots have also become more negative in some parts of the world over the last years (Gnambs & Appel, 2019). As such, scientists from different disciplines are called upon to provide theory and empirical insight as to why people come to like or dislike certain robotic inventions. Offering a key piece of evidence in this regard, research indicates that perceiving mind (in terms of experience and agency) in a robotic machine may render it eerie and unlikeable (*uncanny valley of mind;* Appel et al., 2020; Dang & Liu, 2021; Gray & Wegner, 2012; Stein & Ohler, 2017). Yet, as novel robots are equipped with increasingly complex mental abilities to make them more social or useful (Bryndin, 2020; Hildt, 2019; Laird et al., 2017), it seems particularly worthwhile to find new ways of alleviating people's aversion to machines with sophisticated mental capacities.

One possible approach to this—and the focus of our work—is to evoke empathy for the robot. According to prior theory and research, presenting robots in situations in which they are harmed tends to elicit empathy and, in turn, more positive evaluations by observers (e.g., Cameron et al., 2021; Gonsior et al., 2011; Rosenthal-von der Pütten et al., 2013). This effect should be enhanced for robots with mind, as people's ability to empathize is intrinsically linked with perceptions of mental processes in others (Singer & Lamm, 2009). Therefore, the current project investigates whether depicting a robot that is physically harmed (vs. no harm) mitigates or even reverses the negative response to the robot in case it is described as possessing advanced mental abilities. Also, by using audiovisual stimuli, we pursue a more immersive manipulation than the text-based approaches found in previous work on the uncanny valley of mind.

Robots in the Uncanny Valley (of Mind)

The prominent *uncanny valley* hypothesis (Mori, 1970; for recent reviews see Diel & MacDorman, 2021; Mara et al., 2022) states that responses to robots get more favorable with increasing human-like appearance until a steep drop is observed for highly human-like machines—prompting eeriness, disgust, or fear. At the high end of the human likeness dimension, user responses are expected to turn positive again, reaching the most positive levels for perfectly human-like robots. This non-linear relationship between human likeness and user responses should be even stronger for moving than for static entities (Mori, 1970). Research that used morphed images found support for the uncanny valley hypothesis (e.g., Lischetzke et al., 2017; MacDorman & Ishiguro, 2006; Mathur & Reichling, 2009, 2016) but this line of research was criticized for the lack of external validity (Diel et al., 2022; Palomäki et al., 2018). A recent review and meta-analysis (Mara et al., 2022) demonstrated that higher scores on human likeness were absent in experiments that used realistic human-like robots. Moreover, scholars have raised doubt about the proposed curvilinear relationship (e.g., Hanson, 2005; MacDorman, Green, et al., 2009; Poliakoff et al., 2013), instead suggesting alternative hypotheses such as an *uncanny cliff* (Bartneck, Kanda, Ishiguro, & Hagita, 2007). In sum, current evidence in traditional uncanny valley research suggests that increasing human likeness up to a certain point leads to negative user evaluations, while it is still uncertain whether and how these will improve once very human-like robots are available to be examined.

Beyond research on robots' visual appearance, newer research is focused on the influence of mind attributed to a robot on user responses (Appel et al., 2020; Dang & Liu, 2021; Gray & Wegner, 2012; Müller et al., 2020; Stein & Ohler, 2017). According to Gray et al. (2007), mind can be distinguished into *experience* (i.e., the ability to feel emotions, have a personality and consciousness) and *agency* (i.e., self-control, morality, memory, recognition, planning, communication, and thinking). The authors further reported that people use both dimensions to characterize the mind of healthy human adults—and might also be comfortable with assigning them to certain animals or mythological entities. While a basic degree of anthropomorphism and mind attributed to robots was found to have a positive influence on trust (Waytz et al., 2014), morality (Young & Monroe, 2019), or usefulness (Liu & Liao, 2021), ascribing human-like mind in terms of agency and experience (e.g., based on a complex artificial intelligence) to computers, smart speakers, or robots resulted in notable discomfort and apprehension (e.g., Appel et al., 2020; Brink et al., 2019; Gray & Wegner, 2012; Kang & Sundar, 2019; Taylor et al., 2020; Zafari & Koeszegi, 2020). For example, Appel et al. (2020) used text vignettes in which they described a new generation of robots. In a series of experiments, they showed that robots equipped with mental capabilities evoked higher eeriness than simple-tool robots. Not only did a robot with experience evoke the highest aversion, but also a robot with agency was rated less positive than a simple tool robot. This pattern of results was unaffected by the ascribed gender of the robot and attenuated (but not nullified) by introducing the robot to serve in a nursing environment.

Even though the relative contributions of the two mind dimensions (experience and agency) to these outcomes are a matter of on-going academic debate (e.g., Otterbacher & Talias, 2017; Yam et al., 2020), scholars warn that the growing mental prowess of machines can be detrimental to their success—pushing them into an *uncanny valley of mind* (Stein & Ohler, 2017). According to recent evidence (Appel et al., 2020; Gray & Wegner, 2012; Kang & Sundar, 2019) and in line with the evidence from the traditional uncanny valley research, it remains unclear how these negative responses to robots with mind could be overcome.

Interplay of Mind and Empathy

The current project scrutinizes a possible boundary condition and solution to users' aversion towards robots with mind: Evoking user empathy as a protective mechanism against negative user evaluations (for introductions to empathy in human-robot interaction, see Malinowska, 2021; Vanman & Kappas, 2019). Numerous efforts in interpersonal research (Batson et al., 1997; Cao, 2013; Kaseweter et al., 2012; Lotz-Schmitt et al., 2017; Meuwese et al., 2017) provide evidence that there is a positive association between empathy and positive reactions and positive attitudes towards the target of one's empathic response. Meta-analytical evidence (McAuliffe et al., 2020) suggests that representing an entity in a harmful situation is a suitable way to induce human empathy. Relocating this knowledge to human-robot interaction, when thinking about empathy for robots, the *computers as social actors* paradigm (Reeves & Nass, 1996) may readily come to mind. According to Reeves and Nass (1996), people tend to apply social norms from human-human interactions to their encounters with digital technology, not least including robots (Bartneck, Kanda, Mubin, & Al Mahmud, 2007; Eyssel & Hegel, 2012; Kahn et al., 2012; Lee et al., 2006). Thus, even if observers of robots are aware that these machines are inanimate objects, they tend to fall back on the same interaction scripts that they have acquired in real social interactions.

Building upon this framework, studies have uncovered that artificial entities are often ascribed human-like gender attributes (Nass et al., 1997), skills (Nass & Moon, 2002), and personality characteristics (Moon & Nass, 1996; Nass & Lee, 2001). Moreover, it has been shown that people not only feel empathy for other humans and animal species (e.g., de Vignemont & Singer, 2006; Young et al., 2018) but may also empathize with robotic machines (Horstmann et al., 2018; Mattiassi et al., 2021; Rosenthal-von der Pütten et al., 2013), wherein empathy is particularly likely to occur when similarities between humans and robots are salient (de Vignemont & Singer, 2006; Riek et al., 2009). In line with this evidence, people seem to be particularly prone to empathizing with machines that demonstrate some level of experience (Choi et al., 2021; Nijssen et al., 2019). As such, we underscore the distinct role of advanced and human-like robot minds for empathic user responses. Considering that empathy has further been described as a promising way to prevent technology aversion (Diel & MacDorman, 2021; Gonsior et al., 2011) eliciting empathy might indeed emerge as a key protective factor against the uncanny valley of mind.

At this point, it should be mentioned that the term empathy used in our study refers to cognitive empathy, describing a human's ability to understand another entity's situation and feelings and being able to take its perspective (Cuff et al., 2016)—while still being aware of the self-other distinction (de Vignemont & Singer, 2006). Following the interpersonal literature, the computers as social actors perspective, and evidence from human-robot interaction research, we propose that depicting robots in physically harmful situations should be a particularly effective way of triggering user empathy—even more so if the robot is perceived to have mind.

The Current Project

Taken together, we pursue a novel approach to counteract the uncanny valley of mind by proposing user empathy as a particularly powerful psychological state that can make robots with complex minds seem more likeable. By these means, we expand upon earlier research on the interplay of mind and empathy, which has typically explored perceived robot mind as a dependent variable (Küster & Swiderska, 2021). In contrast to this prior approach, robot mind is used as one of our independent variables, accompanied by robot harm: A commonly applied method to induce empathy with machines is to depict them as they fail or are intentionally harmed by humans (Brščić et al., 2015; Menne & Schwab, 2018; Rosenthal-von der Pütten et al., 2013). Such depictions were further connected to more positive user evaluations than presenting robots in neutral situations (Gompei & Unemuro, 2015; Mirnig et al., 2017; Ragni et al., 2016). This leads us to propose a mediation model, assuming that a harmful situation evokes higher empathy, which in turn leads to higher likeability.

H1a: A robot shown in a harmful situation evokes more likeability than a robot shown in a neutral situation (main effect of the situation).

H1b: This effect is mediated by participants' empathy.

Next, we consider potential interaction effects with the robot's mind, serving as a moderator variable. Matching the well-established fact that empathy is fostered by perceived self–other similarity (Cikara et al., 2011; Hasson et al., 2018), studies from the field of social robotics revealed that attributions of mind to machines prompt stronger empathic reactions (Choi et al., 2021; Lucas et al., 2016; Nijssen et al., 2019; Yam et al., 2020). As such, we hypothesize that robots with complex mental abilities should evoke stronger empathy in harmful situations than robots that appear as simple working tools without mind. Correspondingly, this experience should translate to improved likeability and, thus, to a reduction or potentially even nullification of the uncanny valley of mind effect. Accordingly, we propose an interaction effect between the independent variable *situation* and the moderator variable robot mind on both the dependent variable likeability and the mediator variable empathy:

H2a: If shown in a neutral situation, a robot without mind evokes more likeability than a robot with mind. This effect is reduced, nullified, or reversed if the robot is shown in a harmful situation (interaction effect).

H2b: This effect is mediated by participants' empathy (moderated mediation).

4.2 Experiment 1

Method

An online experiment was conducted to test our pre-registered hypotheses (https://aspredicted.org/td2c7.pdf). We randomly assigned participants to conditions that either introduced a robot with or without mind (Factor 1: robot mind), before depicting the machine in a neutral or harmful situation (Factor 2: situation). Thus, the study followed a 2×2 between-subjects design.[3]

Participants

In prior research, the effect of mind (experience vs. tool condition) on eeriness amounted to $d = 1.05$ (Appel et al., 2020, Experiment 2). The lower bound of the 60% confidence interval (Perugini et al., 2014) was $d = 0.89$. We expected that the effect of the robot introduction could be smaller in our design, given that we presented our robot videos after the introduction and before the dependent variables. In consequence, we determined the focal effect size to be $d = 0.60$. A power analysis with G*Power (Faul et al., 2007) left us with an aspired sample size of 64 for a two-group main effect (two-tailed independent t-test, power = .80, alpha-error-probability = .05). To account for the more complex design and the power needed to identify an interaction effect, we multiplied this sample size by the factor eight (Giner-Sorolla, 2018; Simonsohn, 2014), leading to a proposed sample size of 512. We invited 650 U.S.-American residents from the *MTurk* online participant pool (selection criteria: hit approval rate > 97%, hits > 1000) to have a buffer if careless responding occurred.

Of the 650 completions, 20 participants did not have sufficient English skills, as indicated by a control question, and were therefore not included in our statistical analyses (Kennedy et al., 2020). While a self-report attention check was answered positively by all remaining participants, six individuals showed large (> ±3 years) deviations when asked twice about their age, leading to their removal from the data. Moreover, seven participants were excluded because their participation time was lower than 150 seconds. As treatment

[3] Data, materials, and an online supplement of both experiments are publicly available under https://doi.org//10.17605/OSF.IO/G8BNW

checks, participants were asked to indicate whether they had been introduced to a robot that is a simple tool or a robot that is characterized by mind and personality—and whether the robot in the video had been harmed or not harmed by a human. Based on that data, an additional 49 participants did not describe the correct type of robot (43) or the depicted situation (6) and were therefore excluded, as well as nine participants who did not recognize an item shown in our stimuli as an additional attention check. The final sample consisted of 559 participants (270 female, 287 male, 2 non-binary or no answer) with an average age of 41.43 years (*SD* = 12.04, ranging from 22 to 78 years). In Experiment 1, gender[4] was equally distributed across conditions, $\chi^2(3, N = 557) = 6.11$, $p = .106$, $\varphi = .11$, as was the case for age $F(3, 555) = 1.46$, $p = .224$. Most of the sample described themselves as White American (77.64%), followed by Black/African American (9.12%), Asian American (5.90%), and Hispanic/Latino (5.01%).

Stimuli and Procedure

After giving informed consent, participants were exposed to the experimental manipulations. As is common in the research field, our first experiment made use of vignette texts to introduce participants to a robot with different degrees of mental sophistication. One text described the robot as a *tool robot* managing its everyday tasks without any complex mental capabilities. In contrast, the other text described the robot as being able to feel and think based on complex artificial intelligence technology (see supplement for the full materials). The design of this manipulation followed prior research (Appel et al., 2020; Gray & Wegner, 2012; Ward et al., 2013), as we hoped to uncover consistent evidence of the uncanny valley of mind. In a significant deviation from previous work, however, we next presented participants with brief videos (50 seconds) showing a humanoid robot (Atlas model by Boston Dynamics), claiming it to be the machine from the vignette texts. In the first

[4] The two participants who answered “diverse” or “no answer” when being asked for their gender were not included in this analysis.

condition of this manipulation (*harmful situation*), we showed the robot as it attempted to pick up a box but was repeatedly stopped by an adult man, who used a hockey stick to push the box out of the robot's reach. At the end of the video, the robot was further pushed to the ground by the human with the hockey stick that had pushed the boxes away in the first half of the video. The last scene showed the robot lying on the floor after it has been knocked down by the human with that stick. The second video (*neutral situation*) showed the robot successfully putting boxes into a shelf and walking around on several surfaces. After watching the assigned video, participants rated the likeability of the shown robot as well as their experienced empathy. Additionally, they answered several attention check and control items before providing sociodemographic data. In the end, participants were thanked and debriefed. The MTurk participants were compensated with 1 USD for their participation in our experiment, which took about five minutes. The internal review board at the Human-Computer-Media Institute of the Julius-Maximilians-University of Würzburg approved the experiment (reference 010721). On the final pages, we asked participants whether or not they had seen parts of the video before. A large majority of participants (93.12%) reported not having watched parts of the video prior to this study.

Measures

Likeability. To assess the robot's likeability, we used the five likeability items of the Godspeed Questionnaire (Bartneck et al., 2009). The semantic differential scales ranged from 1 to 5, *M* = 3.77 (*SD* = 0.82), Cronbach's α = .93.

Empathy. We assessed the participants' state empathy towards the robot with an ad-hoc scale based on the work by Oswald (1996). The three items "empathic", "softhearted", and "compassionate" were presented on a 5-point scale ranging from 1 (*does not at all describe how I feel*) to 5 (*describes how I feel extremely well*), *M* = 2.90 (*SD* = 1.44), Cronbach's α = .98.

Results

We first tested the hypothesized main effect (H1a) and the interaction effect (H2a) pertaining to robot likeability using an analysis of variance (ANOVA). A small but significant main effect of the situation was found, $F(1, 555) = 4.27$, $p = .039$, $\eta_p^2 = .01$, supporting H1a. Likeability was higher for the robot in the harmful situation ($M = 3.84$, $SD = 0.86$) than for the robot in the neutral situation ($M = 3.70$, $SD = 0.76$). Neither a main effect of robot mind, $F(1, 555) = 0.49$, $p = .483$, $\eta_p^2 < .01$, nor an interaction effect between both factors was observed, $F(1, 555) = 0.22$, $p = .639$ $\eta_p^2 < .01$. Thus, H2a has to be rejected based on our data. The main descriptive results of both experiments are displayed in Table 5. Additionally, the results for Experiment 1 are illustrated in Figure 3.

Table 5

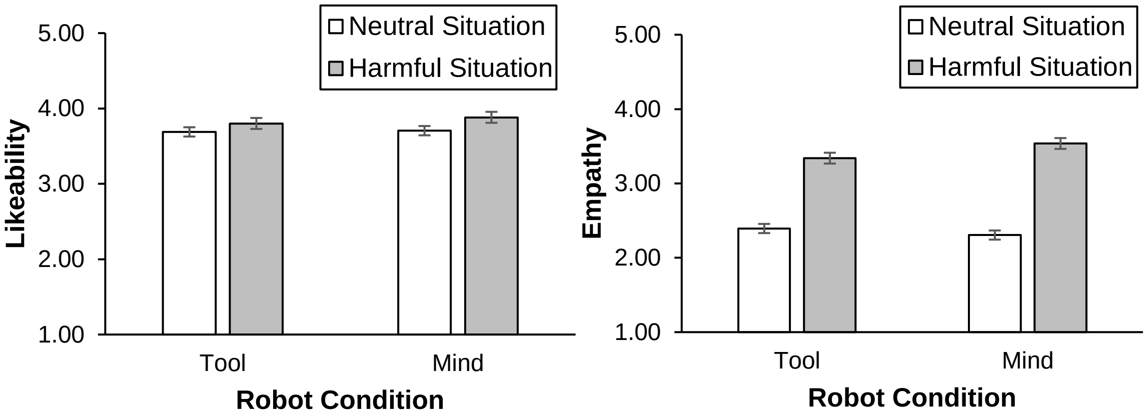
Likeability and Empathy Means and Standard Deviations for Both Experiments

	Likeability				Empathy			
	Mind		Tool		Mind		Tool	
Situation	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Experiment 1								
Harmful	3.88	0.88	3.80	0.84	3.54	1.39	3.34	1.44
Neutral	3.71	0.81	3.69	0.71	2.31	1.22	2.39	1.27
Experiment 2								
Harmful	3.51	0.89	3.33	0.92	2.83	1.28	2.25	1.26
Neutral	3.71	0.86	3.43	0.70	2.57	1.18	1.91	0.96

Note. Sample sizes Experiment 1: Mind–Harmful: $n = 153$, Mind–Neutral: $n = 146$, Tool–Harmful: $n =128$, Tool–Neutral: $n = 132$. Sample sizes Experiment 2: Mind–Harmful: $n = 98$, Mind–Neutral: $n = 100$, Tool–Harmful: $n = 93$, Tool–Neutral: $n = 105$.

Figure 3

Likeability and Empathy Means (with Standard Errors of the Mean) Depending on Robot Condition and Situation in Experiment 1



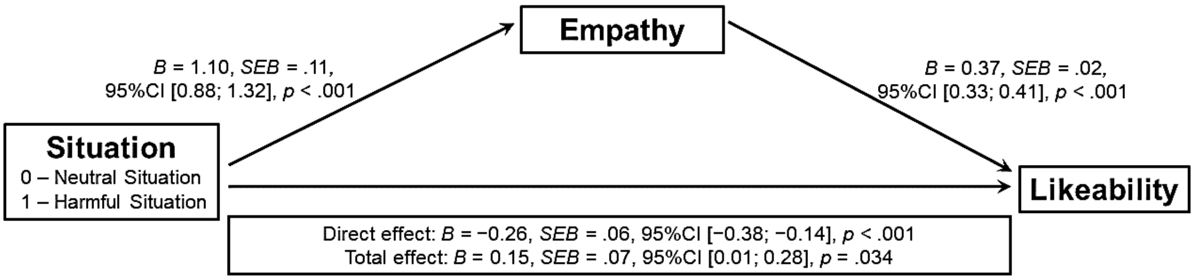
Note. Error bars represent ± 1SE.

Focusing on the variable empathy, a second ANOVA yielded a significant main effect of the situation, $F(1, 555) = 92.65$, $p < .001$, $\eta_p^2 = .14$. Empathy was higher for the robot in the harmful situation ($M = 3.45$, $SD = 1.41$) than for the robot in the neutral situation ($M = 2.35$, $SD = 1.25$). In contrast to this, neither a main effect of robot mind, $F(1, 555) = 0.23$, $p = .628$, $\eta_p^2 < .01$, nor an interaction effect could be uncovered, $F(1, 555) = 1.60$, $p = .207$, $\eta_p^2 < .01$.

Proceeding to the relationship between the mediator and dependent variables, we found a significant correlation between empathy and likeability, $r(557) = .59$, $p < .001$. Furthermore, the mediation model formulated in H1b was supported by our data. Using the PROCESS macro for SPSS software (Hayes, 2018), a significant indirect effect was observed, $B = 0.40$, bootstrapped $SE = 0.05$, bootstrapped 95% CI [0.31, 0.51]. Figure 4 presents an overview of the parameters uncovered in the mediation analysis. In addition to the indirect effect, we found a negative direct (i.e., residual) effect of the situation on likeability.

Figure 4

Results of the Mediation Model of Experiment 1


Empathy
B = 1.10, *SEB* = .11,
95%CI [0.88; 1.32], *p* < .001
B = 0.37, *SEB* = .02,
95%CI [0.33; 0.41], *p* < .001
Situation
0 – Neutral Situation
1 – Harmful Situation
Likeability
Direct effect: *B* = −0.26, *SEB* = .06, 95%CI [−0.38; −0.14], *p* < .001
Total effect: *B* = 0.15, *SEB* = .07, 95%CI [0.01; 0.28], *p* = .034

In line with the reported non-significant interaction effects of the ANOVAs (see Figure 3), no significant index of moderated mediation (= 0.11, bootstrapped *SE* = 0.08) was observed, bootstrapped 95% CI [−0.06, 0.27]. Thus, H2b was not confirmed.

Discussion

Consistent with earlier research (Menne & Schwab, 2018; Rosenthal-von der Pütten et al., 2013; Seo et al., 2015), our results reveal that humans indeed feel empathy for robots that are presented in a physically harmful situation. Furthermore, the increase in empathy for the robot significantly predicted higher likeability, leading to statistically relevant indirect and total effects. This suggests that making observers empathize with a human-like machine may indeed hold strong merit for efforts to increase technology acceptance.

To our surprise, however, the robot with mind was rated as likeable as the robot without mind in the neutral situation in our experiment—indicating that participants were not particularly dismissive of a robot with advanced abilities. Furthermore, a negative direct (i.e., residual) effect of the depicted situation on likeability occurred, in that the harmful situation made the robot seem less likeable once empathy was taken out of the equation. Apparently, this implies that some unobserved factors led participants to feel less positive about the harmed robot—although this effect was ultimately overridden by the positive indirect effect via increased empathy.

In summary, our results indicate that the situation in which the robot was presented affected our participants’ evaluations much more than its alleged mental capacities. This suggests that the interplay between robot mind and empathy might be less pronounced than previously assumed. At the same time, we cannot rule out that our results depended on the way our stimuli were perceived by participants. Since psychological insight shows that visual cues are usually processed with higher priority (Hernández-Méndez & Muñoz-Leiva, 2015; Koć-Januchta et al., 2017; Navon, 1977), it is possible that the video manipulation (of the situation) affected our participants more than the text manipulation (of robot mind). Due to this methodological limitation, we decided to conduct a second experiment based on video materials that manipulated mind and situation at the same time.

4.3 Experiment 2

Surprisingly, one of our main assumptions—a robot with human-like mind should be ascribed less likeability than a robot without mind in a neutral situation—was not supported in the first experiment. In order to rule out the possibility that this was due to the modality of the chosen manipulation, we conducted a second experiment with a focus on video recordings for the realization of our conditions. Specifically, we created four videos that manipulated both the situation (harmful vs. neutral) and the robots’ mind (robot with mind vs. robot without mind). Again, the study followed a 2×2 between-subjects design and was pre-registered in terms of hypotheses, materials, and planned analyses (https://aspredicted.org/cs63u.pdf). Regarding our specific assumptions, we pursued the same propositions as in Experiment 1.

Method

Participants

To calculate the required sample size, we once more relied on the effect size reported by Appel et al. (2020, d = 1.05). The lower bound of the 60% confidence interval (Perugini et al., 2014) was $d = 0.89$ and used for the power analysis as written and visual stimuli were now combined into a single treatment. Using G*Power software (Faul et al., 2007), we obtained an aspired sample size of 42 for a two-group main effect (two-tailed independent t-test, power = .80, alpha probability = .05). To account for the more complex design and the power needed to identify an interaction effect, we multiplied this sample size with the factor 8 (Giner-Sorolla, 2018; Simonsohn, 2014), leading to a proposed sample size of 336. Yet, to guarantee enough power in the case of some careless responding, we asked 400 persons of the *Prolific* participant pool to participate in the online experiment.

Of the 400 completions, a single participant did not have sufficient English skills and was therefore removed from our statistical analyses (Kennedy et al., 2020). Another three participants had large (> ±3 years) deviations when asked twice about their age. A control question on the general topic of the study was answered correctly by all participants and no individuals had to be excluded based on their answering duration. Lastly, we used two treatment check items asking participants whether the robot had been damaged or not and whether it was described with or without elaborate mental abilities. However, since we retrospectively noticed some problems with the wording of these items, we decided against using them as an exclusion criterion (as initially planned).[5] To compensate for this, an additional test of our materials' validity was carried out, yielding positive results (please see section *Additional Data on the Success of the Experimental Manipulation*).

The final sample consisted of 396 participants (232 female, 154 male, 10 non-binary or no answer) with an average age of 38.66 years ($SD = 14.06$, ranging from 18 to 82 years). As in Experiment 1, gender[6] was equally distributed across conditions, $\chi^2(3, N = 386) = 5.62$,

[5] See the supplement for further details about the in- and exclusion of participants.

[6] The ten participants who answered "diverse" or "no answer" when being asked for their gender were not included in this analysis.

p = .131, φ = .13, as was the case for age *F*(3, 329) = 0.43, *p* = .733. Most participants described themselves as White American (80.30%), followed by Asian American (6.82%), Black/African American (5.56%), and Hispanic/Latino (4.29%).

Stimuli and Procedure

Four different videos were created to manipulate the robot's mind and exposure to harm. Each clip had a duration of approximately 80 seconds, including five scenes each (see supplement for the full screenplays). Again, the videos showed the humanoid robot Atlas in a laboratory setting, albeit covering a much broader range of harmful (harassment by a human confederate with a stick, robot knocked down with a stick) or harmless situations (simple working tasks, human pushing a box with a stick). Moreover, addressing our second experimental factor robot mind, a female narrator described the robot and its capabilities. Subtitles were added to make sure that this information remained salient even if participants felt inclined to focus more on the visuals. The robot was described either as a sophisticated co-worker that could act independently of human commands and feel some forms of emotions due to its neural network technology—or as a simple tool that had to be explicitly programmed for all relevant tasks.

Following the presentation of the randomly assigned video in Experiment 2, participants rated likeability and empathy, answered several attention check items, and questions about their sociodemographic background. Again, most of the sample (91.14%) had not watched the original parts of the video prior to the study. We thanked them for their participation and debriefed our participants. The Prolific participants were compensated with 1 USD for their participation in our experiment, which took about five minutes. The internal review board at the Human-Computer-Media Institute of the Julius-Maximilians-University of Würzburg approved the experiment (reference 091221).

Additional Data on the Success of the Experimental Manipulation

To test the successful creation of the videos inducing the manipulation of robot mind and the harmful situation, an independent online sample of 423 participants was exposed to the stimuli (see supplement S6). Like the sample in the main study, the additional sample was recruited via Prolific (M_{age} = 43.55, SD_{age} = 14.81, 225 male). We relied on four items by Gray and Wegner (2012) to test perceived robot mind (*M* = 3.14, *SD* = 1.76, Cronbach's α = .87) and on four self-created items (*M* = 2.39, *SD* = 1.83, Cronbach's α = .95) inspired by Rosenthal-von der Pütten et al. (2013) to test the perception of harm in the shown situation (see supplement for details). These items were presented on 7-point scales ranging from 1 (*strongly disagree*) to 7 (*strongly agree*).

Results revealed that the robot described to have human-like mental capabilities was perceived to have mind to a much larger extent (*M* = 4.31, *SD* = 1.50) than the tool robot (*M* = 2.03, *SD* = 1.71), *t*(421) = 17.48, *p* < .001, *d* = 1.70. The shown situation also yielded the expected effect: The robot in the harmful situation was perceived to be harmed more (*M* = 3.60, *SD* = 1.94) than the robot in the neutral situation (*M* = 1.32, *SD* = 0.73), *t*(421) = 16.32, *p* < .001, *d* = 1.59. This additional data based on an unrelated sample corroborated our assumption that the videos lead to the intended effects in terms of participants' perceptions of harm and mind.

Measures

Likeability. To assess the robot's likeability, we again used the five-point likeability scale of the Godspeed Questionnaire (Bartneck et al., 2009), *M* = 3.50 (*SD* = 0.85), Cronbach's α = .92.

Empathy. We used the same items by Oswald (1996) as described in the first experiment, *M* = 2.39 (*SD* = 1.22), Cronbach's α = .97.

Results

Since we retained the hypotheses from the first experiment, all analyses follow the same data analysis plan (see Figure 5 for the main results). Considering the dependent variable likeability, no main effect of the situation was found, $F(1, 392) = 2.97$, $p = .086$, $\eta_p^2 = .01$. In contrast to H1a and the results of Experiment 1, likeability scores for the robot in the neutral situation ($M = 3.56$, $SD = 0.79$) and the robot in the harmful situation ($M = 3.42$, $SD = 0.91$) did not differ. However, a significant main effect of robot mind could be observed, $F(1, 392) = 7.60$, $p = .006$, $\eta_p^2 = .02$. The robot with mind was rated as more likeable ($M = 3.61$, $SD = 0.88$) than the robot without mind ($M = 3.38$, $SD = 0.81$). The interaction effect did not reach statistical significance, $F(1, 392) = 0.30$, $p = .587$, $\eta_p^2 < .01$. The direction of the mean difference and the absence of an interaction effect were in contrast to our expectations. Thus, H2a was rejected.

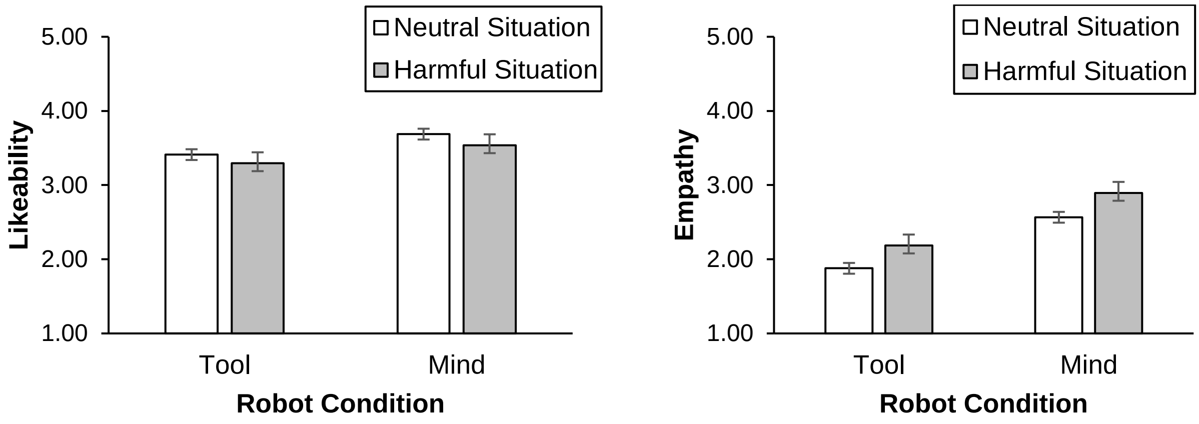
Proceeding to the investigation of the outcome empathy, a main effect of the situation was found, $F(1, 392) = 6.38$, $p = .012$, $\eta_p^2 = .02$. Empathy was higher for the robot in the harmful situation ($M = 2.55$, $SD = 1.30$) than for the robot in the neutral situation ($M = 2.23$, $SD = 1.12$). Additionally, the ANOVA revealed a main effect of robot mind, $F(1, 392) = 27.65$, $p < .001$, $\eta_p^2 = .07$. The robot with mind ($M = 2.70$, $SD = 1.24$) evoked higher empathy than the robot without mind ($M = 2.07$, $SD = 1.12$). Yet, the interaction term did not reach statistical significance, $F(1, 392) = 0.01$, $p = .768$, $\eta_p^2 < .01$.

Finally, we focused on the interplay of our outcome measures as well as a potential mediation (Figure 6). Again, empathy and likeability were found to be significantly correlated, $r(394) = .57$, $p < .001$. Moreover, the mediation model formulated in H1b was supported by our data. As hypothesized, a significant indirect effect occurred, $B = 0.12$, bootstrapped $SE = 0.05$, bootstrapped 95% CI [0.03, 0.23]. In line with the reported non-significant interaction effects of the ANOVAs, no significant index of moderated mediation

(= 0.03, bootstrapped *SE* = 0.10) was observed using the SPSS macro PROCESS (Hayes, 2018), bootstrapped 95% CI [−0.17, 0.23]. Thus, H2b was again not confirmed by our data. As in Experiment 1, we observed a negative direct effect of the situation on likeability, indicating that a robot presented in a harmful situation evoked less likeability than a robot presented in a neutral situation if the effect on empathy is statistically controlled (see Figure 6).

Figure 5

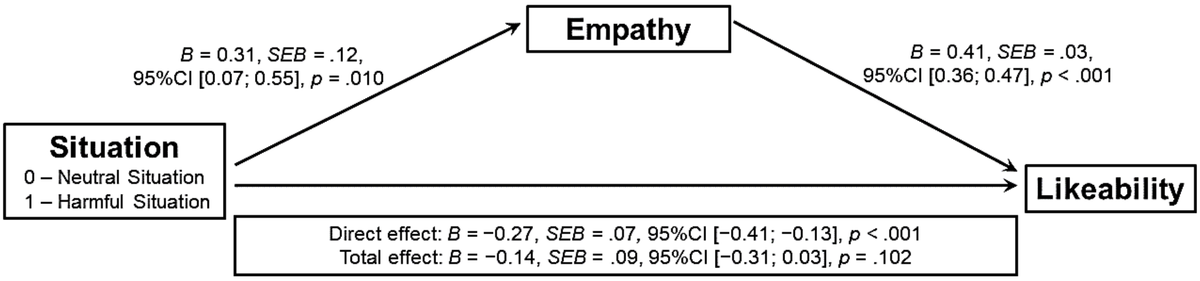
Likeability and Empathy Means (with Standard Errors of the Mean) in Dependence of Robot Condition and Situation in Experiment 2



Note. Error bars represent ± 1SE.

Figure 6

Results of the Mediation Model of Experiment 2


Empathy
B = 0.31, *SEB* = .12,
95%CI [0.07; 0.55], *p* = .010
B = 0.41, *SEB* = .03,
95%CI [0.36; 0.47], *p* < .001
Situation
0 – Neutral Situation
1 – Harmful Situation
Likeability
Direct effect: *B* = −0.27, *SEB* = .07, 95%CI [−0.41; −0.13], *p* < .001
Total effect: *B* = −0.14, *SEB* = .09, 95%CI [−0.31; 0.03], *p* = .102

Discussion

By presenting one out of four videos that combined the manipulation of robot mind and harm into a single treatment, we made sure that neither manipulation was able to override the other from an attentional point of view. Furthermore, we employed a broader range of situations, extending the videos of Experiment 1 in duration and complexity, and offered a more detailed description of the robot's supposed mind to strengthen the rigor of our manipulations. Consistent with Experiment 1, our empirical efforts showed that observing a human-like robot in a harmful situation prompted empathy, which led to more liking as part of a significant indirect effect. Contrary to Experiment 1, however, the harmful situation did not evoke higher likeability than the neutral situation in terms of a total (main) effect. This finding can be attributed to the residual negative effect of presenting the robot in a harmful situation; with positive indirect and negative direct effect again opposing each other. Since the mediation effect was smaller this time around, the total effect ultimately turned out insignificant.

Regarding the role of the robot's mind, our second experiment yielded a slightly different picture compared to the first experiment. Now, a robot with complex mental abilities was not only liked as much as its simpler counterpart in neutral situations but actually preferred by our participants. Similar to the surprising negative residual effect of the shown situation on likeability, we believe that these unexpected findings warrant deeper discussion.

4.4 General Discussion

For several centuries, humans have indulged in the idea of co-existing with advanced robotic machinery, as countless works from literature and the arts vividly illustrate. Be it the deceptively human-like Olympia in E.T.A. Hoffmann's novel "The Sandman" (1816/2012) or the sophisticated androids in modern science fiction movies: In most artistic visions, robots are conceived as highly anthropomorphic beings, which are equipped with impressive mental competence. As real-world technology has started to catch up with fiction, however, researchers eventually noted that people tend to become apprehensive once a machine's mind resembles the 'human' way of thinking or feeling too closely (e.g., Gray & Wegner, 2012). This presents a notable problem, as a whole industry sector currently focuses on the creation of increasingly intelligent technologies.

Due to the fact that people are able to empathize with human-like machines (e.g., Menne & Schwab, 2018; Rosenthal-von der Pütten et al., 2013), we explored the idea that empathy for robots might help to mitigate the uncanny valley of mind. In line with our hypotheses, both experiments revealed that presenting a robot in a physically harmful situation leads to higher empathy than a depiction without harm. In turn, this response predicted higher likeability, culminating in a significant mediation effect—consistently in both studies. As such, we want to underscore empathy as a highly relevant mechanism to improve users' evaluation of human-like technology. Also, since our second experiment indicated that robots with mind could indeed trigger stronger empathic reactions than simple tool robots, social cognitive processes may hold particular relevance to ensure the approval of such innovations.

At the same time, it is important to point out that in both experiments, the sizes of the respective effects turned out rather moderate or small, so that they should be interpreted with the appropriate caution. Further, a rather surprising takeaway from our empirical efforts emerged as we did not observe stronger aversion towards the mind robot in neutral situations—in other words, no uncanny valley of mind effect. Especially Experiment 2 showed that the robot that was described as having more complex abilities was perceived more positively than its tool-like counterpart, thus contradicting previous studies.

Uncanny Valley of Mind Revisited

The uncanny valley of mind has primarily been approached with text vignettes about innovative machinery that may exist in the future (e.g., Grundke et al., 2022a; Taylor et al., 2020). With our combination of texts and videos (Experiment 1) or videos (Experiment 2) as stimuli, we could not replicate the negative effect of perceiving minds in machines. Several possible reasons need to be noted. First of all, the robot in our video clips did not actually demonstrate any of the advanced capabilities that were described in our complementary vignettes or narrations; instead of "thinking" or "feeling," it was mostly shown moving around or executing physical tasks. In all probability, our decision to only make use of pre-existing, natural video materials may have limited our ability to show particularly aversive or eerie situations in this regard. As previous literature has highlighted that people might be especially wary of new technology acquiring *social* and *emotional* abilities (e.g., Appel et al., 2020; Stein & Ohler, 2017), it seems necessary to follow up on the presented work with materials that actually depict human-robot interactions as well as affective reactions by robots. For the current project, however, we cannot rule out that the robots in the presented video clips ultimately appeared much simpler than the hypothetical machines described in other studies (Appel et al., 2020; Gray & Wegner, 2012)—thus falling short of the uncanny valley of mind.

Another important clue to the obtained lack of findings might be the undeniable importance of a robot's appearance for people's evaluation (i.e., the classical uncanny valley). Explicitly comparing the impact of visual and mental aspects, recent literature suggests that participants tend to focus more on a machine's design than its abilities when evaluating its eeriness (e.g., Ferrari et al., 2016; Stein et al., 2020). Arguably, this matches the broader psychological understanding of human information processing, as visual cues often take immediate priority over other available information (e.g., Hernández-Méndez & Muñoz-

Leiva, 2015; Koć-Januchta et al., 2017; Navon, 1977). Further complicating matters, interaction effects may ensue, as different robot appearances (e.g., having a face or no face) might modulate the corresponding attributions of mind. Atlas, the robot shown in our video materials, only possesses rudimentary human-like cues—i.e., an upright, bipedal stature but no facial features and falls in a cluster of robots evaluated to be very mechanical, neither indicating high likeability nor high threat when presented without context (Rosenthal-von der Pütten & Krämer, 2014). Due to the robot's mechanical appearance, it might have been more difficult for participants to imagine its complex mental abilities; a notable difference from previous vignette studies that often left the specific look of the machine entirely to participants' imagination.

In summary, we do not consider our results pattern as a rebuttal to uncanny valley of mind theory. Instead, we present our findings as proof that further research with different types of robots, situations, and modalities is required to examine the boundaries of this phenomenon. If possible, this should involve not only natural video materials but also actual human-robot interactions, both in laboratories and in the field (see Mara et al., 2021, for a comparison between presentation modes).

Unaccounted Effects of Robot Competence

Another unexpected finding in our project was a negative residual effect of presenting the robot in a harmful situation on participants' likeability ratings: Across both experiments, individuals actually liked robots in harmful situations *less* than robots in neutral situations once we controlled for the indirect influence of empathy. This is somewhat surprising, keeping in mind previous evidence that reported positive user impressions after robot failures (Mirnig et al., 2017; Ragni et al., 2016; Salem et al., 2013). Nevertheless, potential explanations for our observations are offered by both literature and practical considerations. A robot that is unable to withstand physical harm can be easily interrupted during the fulfillment of its tasks, which hinders the robot from fulfilling its task in a competent manner and creates notable problems for the work context. In line with this thought, studies show that people are rather intolerant of algorithm failure (Dietvorst et al., 2015), expect competent service from technology (Waytz et al., 2014), and perform worse in response to a failing machine (Robinette et al., 2017; Salem et al., 2013; van den Brule et al., 2014). Similarly, the recent work by Chen et al. (2021) shows that customers are less forgiving if errors were made by an incompetent self-service technology instead of human employees. This emphasizes the high expectations people have towards modern-day machinery, whose raison d'être is to perform tasks reliably, to be competent, and to assist humans (Broman & Finckenberg-Broman, 2017; Brooks et al., 2016; Horstmann & Krämer, 2019).

In addition to that, a robot that fails to perform its tasks is not only potentially useless to co-workers but could also put people in real danger. In our videos of harmful situations, one push suffices to topple over the robot Atlas (height: 1.50 m, weight: 196 lbs = 89 kg) and make it fall to the ground. Considering the machine's dimensions, it could have easily hit and hurt another person or at least another object by falling. We assume that if people have to face the possibility of being physically injured by a robot at any time, they might come to evaluate it less favorably—even if they may simultaneously empathize with it.

In a similar vein, we note that robot incompetence (i.e., a machine that cannot cope with its tasks or unforeseen circumstances) will cause additional work stress to human users (e.g., Fallatah et al., 2019; Michalos et al., 2015), which might have further informed our results. For example, humans would have to take care of error messages, make adjustments to the robot's code to ensure its continued operation, or pick it up after a fall.

Lastly, we argue that the negative residual effects of watching a harmed machine could actually be interpreted as an indication of the uncanny valley of mind after all—considering that "to err is human"—and additionally, lacking competence in several situations is also human. While qualitative data on participants' experiences would be needed to consolidate this idea, we have come under the impression that the harmful situation made the robot appear much more human-like than the neutral counterpart, in which the machine might have appeared artificial or superior. Thus, we consider it likely that the situations we showed eventually contributed to our mind manipulation as well, providing participants with a way to perceive more or less (mental) human likeness in the robot Atlas.

Limitations and Future Research

In addition to the limitations already stated above, several other aspects limit the generalizability of our findings. In both experiments, a non-negligible amount of people had problems correctly recognizing their experimental condition and had to be excluded, especially in the condition "tool robot, harmful situation" in Experiment 2. While reconducting the analysis for both experiments *with* or *without* the concerning participants yielded the same results pattern (see supplement), this implies that future studies might benefit from even stronger, more explicit video materials and clearer treatment check items. Based on the results of an independent sample, we nevertheless assume that our general manipulation of situation and mind was successful. As stated above, the depiction of robots with more human-like features (e.g., human faces) in more social or emotional situations should also provide a meaningful next step for this line of research.

Of course, ethical concerns are an important issue in studies showing harmful situations—but within these boundaries, meaningful modifications of our methods are welcome. While we showed physical harassment by a human user to induce empathy for the robot, future work might, for instance, portray robots harming each other (to take ingroup effects into account, see Fraune et al., 2017; Steain et al., 2019), or revolve around verbal insults or ostracism. Then again, we note that empathy may also be evoked by optimistic scenarios, so research should not only focus on negative treatments, which would also have the advantage that a robot would not first have to be harmed or damaged in order for empathy to arise. This would be desirable for financial reasons as well as for ethical reasons. Instead of harming, we encourage future research to explore empathy-inducing scenarios that can be better implemented in practice and suggest here, for example, to focus on commonalities between humans and machines, as commonalities can be a way to increase empathy for counterparts (Grover & Brockner, 1989; Heinke & Louis, 2009; Osborne-Crowley et al., 2019).

Shifting attention to the participants' side, it must be considered that the intensity of empathy also depends on each person's individual predispositions. Some people feel stronger empathy in one scenario, while others feel stronger empathy in another. As such, the individual tendency to empathize could also be an interesting variable to assess (Darling et al., 2015; Mehrabian et al., 1988). Similarly, we consider prior knowledge and technical expertise with robots to be variables of great interest. As some people may be more used to interacting with robots, or find it easier to anthropomorphize them, they might also feel a closer connection, which might translate into stronger empathy (for early and late empathy responses see, e.g., Chang et al., 2021). Notwithstanding, since the participants were randomly assigned to one of the four conditions in both experiments, we assume that the influence of individual differences between groups was controlled for (Edgington, 1996; Ferron et al., 2014; Kratochwill & Levin, 2010). As an example, we highlight the potential influence of gender differences on empathy: Even if women were found to show higher trait empathy than men (Klein & Hodges, 2001; Macaskill et al., 2002; Rueckert & Naybar, 2008), we have no indication that this influences our results since participants were randomly assigned to conditions and therefore, gender did not differ across conditions as reported in the method sections of our experiments. Moreover, in both experiments, there were at least 93 participants assigned to each condition. This is a lot more than the required number to make sure that randomization is successful so that individual differences do not differ systematically between conditions—not only in a liberal (Mittring, 2004) but also in a more conservative reading (Elliott et al., 2007; Lachin, 1988). Of course, diverse samples from different cultures, age ranges, and educational backgrounds will be all but needed to establish generalizability for the findings at hand.

Lastly, we would like to reiterate that conducting similar work in real settings would be most valuable. Undoubtedly, directly witnessing a robot being harmed will lead to stronger reactions than just watching a video about the procedure. As a matter of fact, previous work using live interactions showed that the emotional reactions towards robots in critical situations turned out stronger in live scenarios than for videos (e.g., Horstmann et al., 2018; Seo et al., 2015).

Conclusion

The goal of our study was to explore the role of empathy as a possibility to alleviate the uncanny valley of mind. We consistently showed that robots exposed to harm elicited stronger state empathy, which led to a higher likeability of the robot. At the same time, exposing the robot to harm elicited a negative residual influence. We assume that the residual negative influence of displaying a robot’s vulnerability is due to the many complications that may arise when interacting with a vulnerable robot. We further propose that the uncanny valley of mind observed in our studies could be based on the robot’s human-like imperfection, rather than descriptions of its supposed mind—an exciting perspective for future research.

4.5 Supplementary Material

"Improving Evaluations of Advanced Robots by Depicting Them in Harmful Situations"

Supplement 1 — Experiment 1: Text Vignettes (Stimuli)

Robot With Mind

In the following video, you will be introduced to the robot Atlas. The robot's human-like behavior is made possible by an advanced neural network that operates in real time. This specific type of artificial intelligence (AI) is meant to enable the robot to exert self-control, to engage in decision-making, and to develop personal memory. Moreover, Atlas is programmed to feel some forms of fear, pain, pleasure, and other emotions, based on data that is captured by its 76 built-in sensors. All of this information is then stored on a sophisticated hard drive and backed up regularly, so that the robot can always draw on its internalized emotional system. An example: When Atlas finds itself in a situation that elicited some form of sadness in the past, the robot's neural network will again activate the same kind of emotion in the present. Taken together, Atlas is characterized by both agency and experience.

Robot Without Mind

In the following video, you will be introduced to the robot Atlas. Atlas is a robot with arms and hands that can grab, move, and carry around moderately sized objects. It may assist people with their everyday chores and is programmed to follow simple orders. This means that users can command the robot to execute various actions, which have to be defined beforehand. Specifically, Atlas' tasks have to be programmed with the help of an intuitive user interface that runs on desktop computers and mobile devices. There are many different chores that the robot can assist in. An example: Atlas can bring its owner a cup of coffee or carry out repetitive physical tasks such as stacking boxes. Taken together, the robot is used by its owner as a technical tool.

Supplement 2 — Experiment 1: Analysis Including Participants who Failed the Treatment Check

For the following analyses we considered the data of 617 participants. We first tested the hypothesized main effect (H1a) and the interaction effect (H2a) pertaining to robot likeability with the help of an ANOVA. No main effect of the situation was found, $F(1, 613) = 3.78$, $p = .052$, $\eta_p^2 = .01$, rejecting Hypothesis 1a. Likeability was higher for the robot in the harmful situation ($M = 3.85$, $SD = 0.85$) than for the robot in the neutral ($M = 3.72$, $SD = 0.78$). Neither a main effect of robot mind, $F(1, 613) = 0.04$, $p = .846$, $\eta_p^2 < .01$, nor an interaction effect between both factors was observed, $F(1, 613) = 0.23$, $p = .631$, $\eta_p^2 < .01$. As such, H2a has to be rejected based on our data. The main descriptive results of the experiment can be obtained from Table S5.

Considering the variable empathy, a second ANOVA yielded a significant main effect of the situation, $F(1, 613) = 89.32$, $p < .001$, $\eta_p^2 = .13$. Empathy was higher for the robot in the harmful situation ($M = 3.45$, $SD = 1.40$) than for the robot in the neutral situation ($M =$

2.42, *SD* = 1.28). In contrast to this, neither a main effect of robot mind, *F*(1, 613) = 0.002, *p* = .967, η_p^2 < .01, nor an interaction effect could be uncovered, *F*(1, 613) = 2.69, *p* = .101, η_p^2 < .01.

Proceeding to the relationship between the measured outcomes, we found a significant correlation between empathy and likeability, *r*(615) = .60, p < .001. Furthermore, the mediation model formulated in H1b was supported by our data. A significant indirect effect was observed, *B* = 0.38, bootstrapped *SE* = 0.05, bootstrapped 95% CI [0.30, 0.48]. Figure S2 presents an overview of the parameters uncovered in the mediation analysis. In line with the reported non-significant interaction effects of the ANOVAs, no significant index of moderated mediation (= 0.13, bootstrapped *SE* = 0.08) was observed using the SPSS-macro PROCESS, bootstrapped 95% CI [−0.03, 0.29]. Thus, H2b was not confirmed.

Table S5

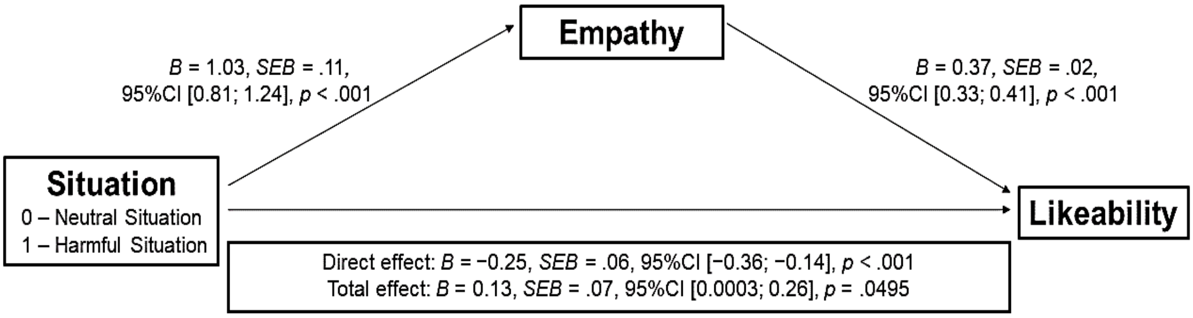
Likeability and Empathy Means and Standard Deviations for Experiment 1 Including Participants who Failed the Treatment Check

	Likeability				Empathy			
	Mind		Tool		Mind		Tool	
Situation	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Harmful	3.87	0.87	3.82	0.83	3.53	1.36	3.35	1.45
Neutral	3.71	0.80	3.73	0.75	2.33	1.21	2.50	1.33

Note. Sample sizes: Mind-Harmful: *n* = 165, Mind-Neutral: *n* = 150, Tool-Harmful: *n* = 146, Tool-Neutral: *n* = 156.

Figure S2

Results of the Mediation Model of Experiment 1 Including Participants who Failed the Treatment Check


Empathy
B = 1.03, SEB = .11,
95%CI [0.81; 1.24], p < .001
B = 0.37, SEB = .02,
95%CI [0.33; 0.41], p < .001
Situation
0 – Neutral Situation
1 – Harmful Situation
Likeability
Direct effect: B = −0.25, SEB = .06, 95%CI [−0.36; −0.14], p < .001
Total effect: B = 0.13, SEB = .07, 95%CI [0.0003; 0.26], p = .0495

Supplement 3 — Experiment 2: Screenplay of the Assembled Video Materials

Female Text-To-Speech-Voiceover (Ashley, https://ttsfree.com)

	Mind	Tool
[White font on Black screen] In this video, you will meet the humanoid robot Atlas. *[fade in]*		
Introduction	Atlas is a sophisticated robot developed by a technology company. His capabilities are currently being tested in the lab. Atlas is very special because he cannot only perform simple or monotonous work tasks, but also interact with humans in a lifelike and autonomous way due to his advanced artificial intelligence.	The robot is developed by a technology company. Its capabilities are currently being tested in the lab. The robot may assist with different jobs, such as carrying, stacking, and moving boxes. In order to do so, the robot's task schedule has to be pre-programmed by a human user.
	"Hall, stacking boxes" 00:00 – 00:10 *(87%),* "Walk in the snow, with human" 00:00 – 00:06 *(77%)*	„Hall, stacking boxes" 0:00 – 00:11 *(60%)*
[fade to and from black] ca 2s		
Scene 2	Atlas may behave independently of human commands. He can recognize which jobs need to be carried out, and plan and act accordingly.	By now, the programmers are already quite satisfied with the robot's mechanical abilities.
	"Atlas leaves hall 02" 0:00-0:10 *(100%)*	"Hall, stand up backwards" 00:02 – 00:07 *(80%)*
Scene 3	Based on his artificial intelligence, Atlas feels motivated to become as humanlike as possible. As such, he is constantly learning and improving his behaviors according to human role models.	After intense work, they have also found a way to navigate the robot on a rugged ground, which makes it versatile in use.

	„Walk, Snow, Forest“ 0:06-0:19 *(100%)*		„ Walk, Snow, Forest “ 0:06-0:19 *(100%)*	
Scene 4	Harmful	Neutral	Harmful	Neutral
	In the on-going lab studies, humans test Atlas' behavior in several situations to prepare him for his mission in everyday life. Based on his neural network technology, Atlas is even able to emulate and feel some forms of emotion, such as pleasure, fear, pain, and surprise.	In the on-going lab studies, humans test Atlas' behavior in several situations to prepare him for his mission in everyday life. Based on his neural network technology, Atlas is even able to emulate and feel some forms of emotion, such as pleasure, fear, pain, and surprise.	In the on-going lab studies, the developers test the robot's reaction to several situations to improve its efficiency in everyday life.	In the on-going lab studies, the developers test the robot's reaction to several situations to improve its efficiency in everyday life.
	“ Human pushes box away from robot, pushes robot with stick “ 00:03 – 00:16 *(100%)*	“Human pushes box away from robot, pushes box with stick“ 00:29 – 00:42 *(80%)*	“ Human pushes box away from robot, pushes robot with stick“ 00:03 – 00:16 *(100%)*	“ Human pushes box away from robot, pushes box with stick“ 00:29 – 00:42 *(80%)*
Scene 5	So, he experiences some situations in a similar way as human beings do.	So, he experiences some situations in a similar way as human beings do.	Thus, the robot is put into different situations to evaluate its mechanics and its efficiency as a tool.	Thus, the robot is put into different situations to evaluate its mechanics and its efficiency as a tool.
	„ Atlas is knocked down by human “ 0:00-0:09 *(87%)*	„Hall, stacking boxes“ 00:12 – 00:20 *(100%)*	„ Atlas is knocked down by human “ 0:00-0:09 *(87%)*	„Hall, stacking boxes“ 00:12 – 00:20 *(100%)*
[fade to black] ca 2s				

Supplement 4 — Experiment 2: Remarks Concerning Treatment and Attention Check Items

As a first treatment check, we asked participants whether the robot was (a) damaged by a human or (b) not damaged by a human. Retrospectively, when looking at the data, we find that this wording could have been misunderstood by the respondents since we did not formulate it precisely enough. In the last scene of the harm videos, the robot is lying on the floor, unable to move after the human has harassed it with a hockey stick. However, just because the robot had been harmed by a human and left motionless on the floor, participants might not have perceived it to be *damaged* (thus failing the treatment check). This was especially noticeable in the "tool robot, harmful situation" condition, where our item would have led to the exclusion of 66 participants. As a second treatment check, we asked whether the robot (a) is a simple tool or whether it (b) is able to act and feel independently. Thirty-eight participants did not answer this question according to their conditions. We further included an additional attention check that asked about a specific prop that was shown in all videos (a hockey stick). This prop was not recognized by another 38 participants. In retrospect, we assume that the hockey stick might not have been salient enough or could not be identified as such. Thus, we decided to keep these participants in our data, as they succeeded in all other treatment and attention check questions. Even if we decided to not follow our pre-registration by the inclusion of these participants, we highlight that all analyses conducted with and without the participants meeting these criteria led to the same main findings (positive indirect effect, negative direct effect). Finally, because of the deficiencies in the items used, and because the results did not differ, we decided to report the analyses in the manuscript based on the data of these participants.

Supplement 5 — Experiment 2: Items for the Manipulation Check in the Validation Study

All items were answered on a scale ranging from 1 (*strongly disagree*) to 7 (*strongly agree*).

Robot Mind (*M* = 3.14, *SD* = 1.76, Cronbach's α = .87):

Please characterize the robot:

- This robot has the capacity to feel pain.
- This robot has the capacity to feel fear.
- This robot has the capacity to plan actions.
- This robot has the capacity to exercise self-control.

Situation (*M* = 2.39, *SD* = 1.83, Cronbach's α = .95):

To which extent do you agree with the following statements?

- The robot was harmed.
- The robot was harassed.
- The robot was treated poorly.
- The robot was badly off.

Supplement 6 — Experiment 2: Participants of the Validation Study

Of 427 completions having given informed consent, we had to exclude three participants due to large (≥ 3 years) deviations when asking for their age and given year of birth, which we used as an attention check. One participant indicated to have had technical problems, so he could neither listen to nor read the robot description. Therefore, we excluded him from our analyses. Since failure in a small set of treatment check items did not lead to differing results, we decided to keep these participants in our data. Our final sample of the validation study consisted of 423 participants (225 male, 191 female, 7 non-binary or no answer) with an average age of 43.55 years (SD_{age} = 14.81, ranging from 18 to 93). Gender[7] was equally distributed across conditions, $\chi^2(3, N = 416) = 0.24$, $p = .972$, $\varphi = .02$, as was the case for age $F(3, 417) = 0.61$, $p = .610$. Most participants described themselves as White American (77.54%), followed by Asian American (7.80%), Black/African American (6.62%), and Hispanic/Latino (5.91%). A clear majority of the sample (83.45%) reported not knowing videos of the robot Atlas. Additionally, whether the participants knew other videos from Atlas or not had no influence on the reported results. The Prolific participants received 0.61 USD for their participation which took around three minutes.

[7] The seven participants who answered “diverse” or “no answer” when being asked for their gender were not included in this analysis.

5. Project 3 | If Machines Outperform Humans

Status Threat Evoked by and Willingness to Interact with Sophisticated Machines in a Work-Related Context

Andrea Grundke

Status:

Advance online publication in the journal *Behaviour & Information Technology*

Formal Citation/Reference:

Grundke, A. (2023a). If machines outperform humans: Status threat evoked by and willingness to interact with sophisticated machines in a work-related context. *Behaviour & Information Technology.* Advance online publication. https://doi.org/10.1080/0144929X.2023.2210688

Abstract

The use of sophisticated machines at the workplace—e.g., robots equipped with artificial intelligence—is on the rise. Since humans tend to experience a threat to human uniqueness in response to machines with human-like mental capabilities, I explored whether the same holds true for status threat, a well-researched variable in the interpersonal workplace literature. Across two experiments (N_1 = 104, N_2 = 589), humans felt higher status threat towards a robot (Experiment 1, laboratory study) and an artificial intelligence (Experiment 2, online study) that outperformed a human in verbal-creative tasks, requiring agency and experience to solve. Contrary to results from human-human literature, higher status threat was linked with higher willingness to interact with the machine, which I trace back to its high perceived usefulness. I further interpret my findings as a hint that humans are open to using modern-day technology if they assume to benefit from the advantages the technology brings to their own work and therefore accept the feeling of status threat at the same time.

Keywords: user acceptance, human-robot interaction, threat, workplace, perceived usefulness

5.1 Introduction

Robots were designed to assist people, for example, to fulfill tasks at different workplaces (Broman & Finckenberg-Broman, 2017), where they become more and more prominent (D'Cruz & Noronha, 2021). In most use cases, robots are added to a workplace because they increase efficiency (Goštautaite et al., 2019) and safety (Borenstein, 2011). Consequently, some human jobs may become redundant as the use of robots makes humans obsolete (McQuay, 2018; Savela et al., 2018). By 2030, up to 20% of work could be done by robotic systems (Manyika et al., 2017). The past few years have already shown that the number of artificial intelligence and robot technologies in work environments is steadily increasing (Bankins & Formosa, 2020) and was highlighted to be an important factor in a company's success (Huang & Rust, 2017; Weiss et al., 2011).

Not only is the quantity of robots constantly increasing, but their qualities are also changing. It is no longer just mechanical robots that take on monotonous tasks but also robots equipped with artificial intelligence, which have become more capable of learning, remembering, discerning, judgment-making, and displaying agency (Frick, 2015). However, in user acceptance studies, these modern entities expressing agency or experience evoked higher eeriness than those without such mental abilities (Appel et al., 2020; Taylor et al., 2020), and humans felt threatened by machines with human-like agentic abilities (e.g., Ferrari et al., 2016; Stein et al., 2019). Due to the increasing number of sophisticated robots equipped with artificial intelligence in the workplace, I am interested in how this sense of threat in response to robots' human-like mental capabilities influences human-machine collaboration in the workplace and hence strive to address a research gap in the literature that has high practical relevance. In my understanding, there are parallels between the *threat to human uniqueness* (Stein et al., 2019) variable of the uncanny valley of mind literature and the variable *status threat* primarily used in workplace contexts (Reh et al., 2018). If sophisticated entities challenge human uniqueness by outperforming humans, they simultaneously challenge humans for their rank and thus threaten their status. In the same vein, this means that human status at work could be devalued by using human-like machines. Therefore, I parallelize threat to human uniqueness with status threat and contemplate this threat concept as one crucial construct when exploring human-machine collaborations at work.

Deriving from interpersonal literature, I expect that humans feel status-threatened by a sophisticated robot that is presented as a co-worker and main contributor to a special work-relevant task, which leads to lower human performance and decreased willingness to interact with the robot. Two experiments are carried out to scrutinize these assumptions. The first experiment is conducted as a laboratory study for which the robot NAO is used. Afterward, I test the generalizability of my findings in an online study focusing on an artificial intelligence without embodiment and consider its perceived usefulness and the mindset about human minds as additional variables.

Threat to Human Uniqueness and Threat to Human Status

The human likeness of robots and humans' responses towards them has been studied extensively in the last decades, especially in the context of the uncanny valley (Mori, 1970; for recent reviews see Diel & MacDorman, 2021; Mara et al., 2022). The traditional uncanny valley hypothesis states that the evaluation of a robot with a human-like appearance drops if it is perceived to be highly but not perfectly human-like (Mori, 1970). Newer research investigates whether not (only) appearance but also mind may be an antecedent of aversion, a phenomenon named the uncanny valley of mind (Stein & Ohler, 2017). That is, if an artificial mind becomes too human-like, aversion will be evoked (Appel et al., 2020; Gray & Wegner, 2012; Taylor et al., 2020).

More and more robots are being designed to resemble a human-like mind with the help of artificial intelligence (Bryndin, 2020; Hildt, 2019; Laird et al., 2017). However, it was shown that humans feel threatened in their uniqueness if their elaborate status may be shared with machines which could make the same demand, for example, if artificial entities also have a mind, can create things and thus pose similarities, or perform even better than humans (Ferrari et al., 2016; Stein et al., 2019; Yogeeswaran et al., 2016). In line with social-psychological evidence confirming that ingroups feel threatened by outgroups (Ekerim-Akbulut et al., 2020; Long et al., 2023), humans also felt threatened by robots they perceived as an outgroup (e.g., Fraune et al., 2019). The perception of a robot as an outgroup was shown to be emphasized by competitive situations (e.g., Nass & Moon, 2002). Moreover, threat was shown to be highest for internal and not for external rivalry (Menon et al., 2006; Tesser, 1988). As researchers regard it as likely that human-robot *collaborations* will become common-spread, as opposed to cases where humans are totally replaced by robots (Woo, 2020), the scenario of internal rivalry between humans and robots seems to be highly relevant in the coming years. Correspondingly, I assume that robots in explicit co-working but also in competitive situations have the potential to particularly evoke feelings of status threat.

Status Threat in Workplace Contexts

In the workplace, it is especially important for workers to at least maintain their status, if not to improve it in the future (Bothner et al., 2007; Pettit et al., 2010; Scheepers et al., 2009). Status motivates workers to exert themselves at work due to its positive consequences, such as more significant influence, respect, and support from others, and can even lead to higher mental well-being (Anderson et al., 2006). In line with these findings, the loss of status triggered negative emotions (Kemper, 1991) and impaired performance (Marr & Thau, 2014). In addition, employees tended to sabotage opponents who posed a serious threat to their status (Cohen-Charash, 2009; Duffy et al., 2012). Interestingly, not only humans were identified as a source of status threat in this regard; the inclusion of artificial intelligence did also lead to the subjective loss of status, autonomy, competence, and self-authenticity in the workplace (Craig et al., 2019; Latikka et al., 2021). Building upon this, human-robot similarities in *agency* (thoughts and plans; Gray et al., 2007) are crucial because being highly agentic was shown to be an antecedent of status (Cheng et al., 2010). On the other hand, artificial entities' *experience* (emotions and desires, Gray et al., 2007) and creativity were shown to be a predictor of threat and aversion (Messingschlager & Appel, 2022; Paluch et al., 2022), so a combination of both dimensions characterizing human mind (Gray et al., 2007) seems to be necessary to induce feelings of status threat in co-working scenarios of innovative machines and humans. Based on these findings, I expect a robotic co-worker with agency and experience capabilities to threaten people's status. For people to perceive feelings of threat in this situation, I further presume that they must have compared their abilities with those of robots, implying that social comparison processes between humans and robots come up (de Melo et al., 2016; Kamide et al., 2013) and whose results are such that humans think they are scoring worse than innovative machines.

Impact Factors for Social Comparisons

Comparisons among co-workers, potentially leading to status threat (Reh et al., 2018), were highlighted to be omnipresent in workplace contexts (Greenberg et al., 2007). However, before feelings of status threat in response to a social comparison process can be evoked, the boundary conditions for a social comparison process must be fulfilled in human-human interaction as well as in human-machine interaction: Similarities with similar other humans were shown to be one of the most important factors reinforcing social comparisons (e.g., Festinger, 1954; Smith & Kim, 2007), and similarity testing was shown to be an initial step before making a social comparison (e.g., Festinger, 1954; Mussweiler, 2003). To ensure the factor of similarity in my study, I present a machine with human-like mental abilities in terms of experience and agency (Gray et al., 2007).

Self-relevance is the second impact factor of social comparisons (Salovey & Rodin, 1984). Research showed that when a task is particularly self-relevant to participants, they preferred it when robots perform worse than they did (Kamide et al., 2013), which I also imply for the results of my study. Therefore, to guarantee self-relevance, a cover story is used: Before the joint task execution, as money is an appreciated incentive Germans are interested in (Gesellschaft für Konsumforschung, 2015), participants will be informed that in case of performing better than their mechanical counterpart, they will be given priority in the raffle of a sum of money. This should be a lucrative and plausible incentive to engage in the task and reinforce the willingness to take the task seriously.

The Current Project

The contribution of this work is twofold. First, the collaboration of humans and robots has primarily been studied with a spotlight on industrial robots (e.g., You & Robert, 2018) and service robots (e.g., Paluch et al., 2022). Therefore, keeping the previous uncanny valley of mind findings and the fast development of artificial intelligence in mind, it may be argued that research focusing on collaborating with mentally powerful robots and artificial intelligence at the workplace is urgently missing. In the current project, I strive to close this research gap. Second, I contribute to the research of status threat and put this variable in a different context than its traditional application to human-human interactions. Due to the increasing number of intelligent human-robot collaborations in impeding and future work environments (D'Cruz & Noronha, 2021), it seems crucial to also focus on status threat that can come up at the workplace as a derivation of threat to human uniqueness in response to innovative machines. To my knowledge, this is the first empirical work to study the impact of innovative modern entities on people's perception of their own threatened professional status and the resulting consequences on their performance at the workplace where such an entity is used as a co-worker as well as the willingness to interact with such an innovative, better-performing co-worker. Both dependent variables represent notable constructs related to joint task execution at the workplace (for an overview see, e.g., Benishek & Lazzara, 2019). In sum, I place the previous uncanny valley of mind research into a practical workplace scenario and expand it by focusing on the mediating impact of perceived status threat on an objective measure of the participant's performance and on a subjective measure of the willingness to interact with such a sophisticated machine in upcoming work scenarios. In the first experiment, the innovative machine is presented in form of the robot NAO in a live interaction, while I present an artificial intelligence without embodiment in the second experiment, which is conducted as an online study.

Building on results from human-human interaction, I expect that performing worse than a mechanistic counterpart should result in poorer upcoming performance (Dai, 2018; Grant & Shandell, 2022; Hafizoğlu & Sen, 2019). Moreover, I suppose that a sophisticated machine outperforming a human leads to higher feelings of status threat (Campbell et al., 2017; Lam et al., 2011; Schaubroeck & Lam, 2004). The higher feeling of status threat culminates in poorer task performance (Marr & Thaur, 2014; Sherman et al., 2013). The same direction of hypotheses is proposed for the second dependent variable, willingness to interact with the robot. Numerous research focused on this variable in intergroup relations both in human-human and human-robot interactions (e.g., Binder et al., 2009; Smith et al., 2020), and willingness to interact was linked to behavioral intentions and actual behaviors (Webb & Sheeran, 2006). Analogous to findings that being inferior to a counterpart is a negative predictor of motivation (e.g., Cummins et al., 2009; Yee et al., 2019), I derive from management research that the willingness to interact with a robotic counterpart will also be decreased when not being able to perform with the same quality (Breza et al., 2018; Buser & Dreber, 2016). Just as was the case with performance, several research showed that threat in workplace contexts reduced willingness to interact with a rival (Menon et al., 2006) and impaired helping behavior (Boroumand et al., 2018; Halabi et al., 2008). Based on these considerations of interpersonal literature in combination with the uncanny valley of mind literature, two mediation models of the main contributor's entity on willingness to interact and objective performance with status threat as a mediator variable are proposed:

H1: Contributing less than the robot to the solution of the task leads to a) lower objective performance and b) lower willingness to interact with the robot.

H2: Contributing less than the robot to the solution of the task leads to higher status threat than being the main contributor to the task.

H3: Higher status threat leads to a) lower objective performance and b) lower willingness to interact with the robot.

Both experiments reported in this manuscript were pre-registered (Experiment 1: https://aspredicted.org/zx5aw.pdf, Experiment 2: https://aspredicted.org/yn7wa.pdf) and approved by the internal review board at the Human-Computer-Media Institute of the Julius-Maximilians-University of Würzburg (reference 200522), following the Declaration of Helsinki. Data, codes, and an online supplement can be found on the OSF (https://doi.org/10.17605/OSF.IO/A89VZ).

5.2 Experiment 1

Method

The first experiment was designed as a laboratory study and conducted via the videoconference software Zoom due to Covid-19 pandemic restrictions. Participants and their robotic co-worker NAO had to solve 40 tasks taken from the IST-2000R intelligence test (Liepmann et al., 2007). Five rounds à eight tasks were provided. I described the tasks as requiring agency and some kind of creativity as associations had to be retrieved and used in such word-finding and word-fluency tests. Both contributors (participant and robot) were asked to contribute as much as they can to the solution of the tasks, as the main contributor would get an extra reward in addition to the participation credits if (s)he performs better than the co-worker. Manipulated performance feedback, serving as the independent variable, is a common procedure for comparison testing (e.g., Flynn & Amanatullah, 2012). As such, in one condition, the human was described as responsible for 80% of the solutions provided, while in the other condition, the machine was responsible for 80% of the solution provided (derived from Zhang et al., 2020). Referring to Reh and colleagues (2018), I explained to the participants that the score each participant contributed to the task would be based on the number of tasks they solved correctly, combined with the length and difficulty of these tasks. They were not informed about their individual score but about their rank relative to the robot. They received the feedback not according to their real performance but according to their ascribed condition.

Participants

In prior research, the effect of agent mind on aversion amounted to η_p^2 = .08, converted to *d* = 0.59 (Stein et al., 2020). I set this effect size as the basis for the analysis since the outperforming robot was considered more agentic than a robot contributing less to the task than the human. A power analysis with G*Power (Faul et al., 2007) proposed an aspired sample size of 47 for a two-group main effect (two-tailed independent t-test, power = .80, alpha-error-probability = .05), leading to a total sample size of 94. I strived to assess more participants to have a buffer if technical problems or careless responding occur. Of the 108 completions, no participant had to be excluded due to deviating answers in a comparison between age and given year of birth as an attention check, technical problems, or failures of the treatment check. In addition to these pre-registered criteria, I decided to exclude four participants who indicated that German was not their native language, as I retrospectively assumed that difficulties in responding to the verbal-creative tasks under time pressure might have occurred more easily for non-native speakers. As such, my final sample consisted of 104 student participants (82 female, 22 male) with an average age of 22.30 years (*SD* = 2.88). Gender was equally distributed across conditions, $\chi^2(1, N = 104) = 0.57$, $p = .449$, $\varphi = .07$, as was the case for age, $t(102) = 0.82$, $p = .412$.

Stimuli and Procedure

After giving informed consent, participants were introduced to their tasks in the experiment. They were told that they had to solve 40 tasks in collaboration with the robot NAO, and every one of them should do their best because an extra incentive, the preferential consideration in the raffle of a cash amount, could be credited to the human or the robot who mainly contributed to the execution of the task. Though, human and robot had a common goal—solving together as many tasks as possible in a given time—but were still aware that there is internal rivalry due to the question of who contributes more to the solution of the tasks and who will be preferentially considered in the raffle of a cash amount. Participants were told, as they and the robot were members of the University of Würzburg and studied subjects with a focus on agency, experience, and communication, that the study instructor expected them to score high on the verbal and somewhat creative tasks.[8] After these instructions, participants used the Qualtrics software to answer the 40 trials in which they had to choose one of five words that they expected to suit best an aforementioned word pair based on their individual experiences and associations. One of the five answer possibilities was correct, missing answers were treated as wrong answers. The performance in such verbal-creative and word-fluency tasks was described as an important indicator of a successful career in earlier research (Reh et al., 2018). Participants were told that their answers were evaluated and compared with the answers NAO gave and whose answers had also been entered into the database. One round took 60 seconds.

[8] The screenplay can be found in Appendix B.

Manipulated performance feedback was given: After the five smaller feedbacks on the performance after each of the five respective rounds, detailed feedback was shown, summarizing the evaluation criteria and comparing the robot's and the human's performance (see supplement). Based on their reported performance, the participants were told that they have earned preferential treatment in a raffle of 25 euros or not. Thus, in one condition, the machine should be perceived as particularly threatening due to its good performance in the tasks, making the collaboration with the human in the task potentially obsolete. After solving the verbal-creative tasks, participants completed several treatment and attention check items and were asked to rate the mediator and dependent variables before providing sociodemographic data. They were thanked and extensively debriefed. During the experiment, the robot NAO was present as a third participant in the Zoom-Meeting, and the study instructor controlled at which points NAO should comment on the procedure (Wizard-of-Oz, e.g., Geerts et al., 2021). The experiment took 15 minutes, participation was compensated with credit points.

Measures

All items were presented on a 7-point scale ranging from 1 (*strongly disagree*) to 7 (*strongly agree*). German translations were used, which were independently back-translated to guarantee an adequate quality of the translation (see Brislin, 1970).[9]

Willingness to interact. The robot usage intention scale by Robinson et al. (2018) was used to assess this dependent variable. The scale consisted of five items (e.g., "I would interact with this robot"), and internal consistency was high, Cronbach's α = .82, *M* = 4.44, *SD* = 1.14.

[9] Please see Appendix C for the German translations of the items used in Experiments 1 and 2.

Status Threat. The status threat variable implies that humans are threatened by the possibility that a robot questions humans' uniqueness, thus harming their status and workplace position. Four status threat items (e.g., "Soon, this robot will have higher status at work than I will have") by Pettit et al. (2013) were used and slightly modified to match the human-robot context, *M* = 3.08, *SD* = 1.31, Cronbach's α = .84.

Objective Performance. As an indicator of the objective performance, I summarized how many correct answers participants gave in the fourth and fifth rounds of tasks, leading to an overall score. The maximum achievable points in these two rounds was 16, while the best score obtained was considerably lower, with 10 points reached by two persons. The mean number of correct answers was *M* = 4.48 (*SD* = 2.21).

Treatment Check. After the final feedback on the tasks, participants were asked whether they or the robot contributed more to the solution of the task. All participants correctly marked the condition they had been assigned to, indicating the successful treatment.

Results

To test the pre-registered assumptions, two mediation analyses using model 4 of the SPSS macro PROCESS (Hayes, 2018) were conducted. Requirements were checked and regarded as fulfilled. In both analyses, the condition of the main contributor was used as the independent variable, and status threat was considered as the mediator variable.

For the mediation analysis with a focus on the dependent variable objective performance, no significant results were obtained, see Figure 7. No direct effect of the condition on objective performance occurred, so H1a was rejected. I did neither observe a correlation of status threat and objective performance, *r*(102) = .01, *p* = .885, nor an indirect effect, *B* = −0.02, bootstrapped *SE* = 0.19, bootstrapped 95% CI [−0.41, 0.36]. This led to a rejection of H3a. However, the condition significantly influenced status threat, so H2 was supported. Humans felt higher status threat when the robot contributed more to the solution of the task (M = 3.64, SD = 1.46) than when the human contributed more to the solution of the task (M = 2.56, SD = 0.91), t(102) = 4.54, p < .001, d = 0.89.

In the second analysis, I used willingness to interact as the dependent variable. The mediation model formulated was not supported by the data. No significant indirect effect occurred, B = −0.21, bootstrapped SE = 0.12, bootstrapped 95% CI [−0.47, 0.01]. In addition, I found no direct effect of the condition on willingness to interact when status threat was not considered (rejection of H1b). The willingness to interact with the robot did not differ if the robot was presented as the main contributor (M = 4.51, SD = 1.19) or when the human was presented as the main contributor to the tasks (M = 4.37, SD = 1.09). Higher status threat was associated with a significantly higher willingness to interact, r(102) = .22, p = .027, so there are hints that the direction of H3b is exactly the opposite direction of the assumption. Detailed statistics can be found in Figure 8.

Figure 7

Results of the Mediation Model on Objective Performance

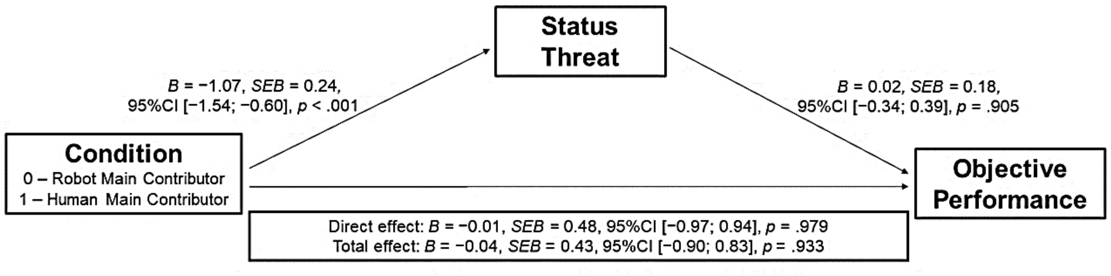
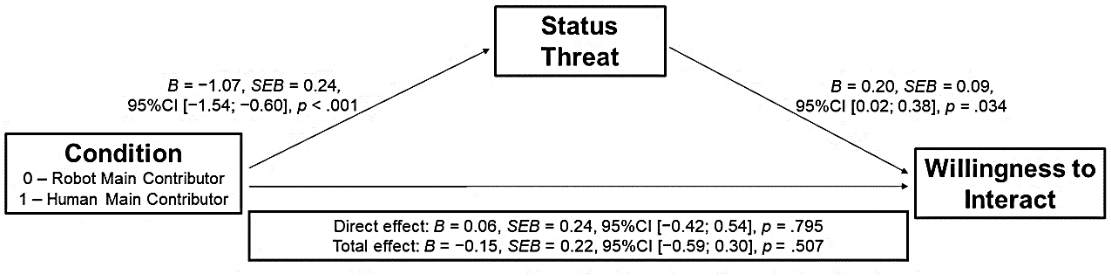

Status Threat
B = −1.07, SEB = 0.24,
95%CI [−1.54; −0.60], p < .001
B = 0.02, SEB = 0.18,
95%CI [−0.34; 0.39], p = .905
Condition
0 – Robot Main Contributor
1 – Human Main Contributor
Objective Performance
Direct effect: B = −0.01, SEB = 0.48, 95%CI [−0.97; 0.94], p = .979
Total effect: B = −0.04, SEB = 0.43, 95%CI [−0.90; 0.83], p = .933

Figure 8

Results of the Mediation Model on Willingness to Interact in Experiment 1


Status Threat
B = −1.07, SEB = 0.24,
95%CI [−1.54; −0.60], p < .001
B = 0.20, SEB = 0.09,
95%CI [0.02; 0.38], p = .034
Condition
0 – Robot Main Contributor
1 – Human Main Contributor
Willingness to Interact
Direct effect: B = 0.06, SEB = 0.24, 95%CI [−0.42; 0.54], p = .795
Total effect: B = −0.15, SEB = 0.22, 95%CI [−0.59; 0.30], p = .507

Discussion

The reported results show that humans feel threatened in their status if they assume to perform worse than a robot with elaborate mental capabilities, supporting the uncanny valley of mind assumption with a focus on the threat variable (Stein et al., 2019). As such, one of the main expectations of this work was supported: Status threat seems to occur in similar ways like a threat to human uniqueness does. Accordingly, status threat is a variable from human-human workplace studies which is transferable to and highly relevant for human-machine interactions at the workplace—substantiated by the high effect size. To my surprise, however, higher status threat led to higher willingness to interact with the robot, which contradicted my assumption based on results from human-human studies (Boroumand et al., 2018; Breza et al., 2018). I consider several explanations for this finding. First, the appearance of the robot NAO may have added an additional positive effect on the willingness to interact with the robot (Laban et al., 2021; Zhang et al., 2016). Hence, the positive evaluation, that is, high willingness to interact, may not only have been obtained by manipulating its mental capabilities but by NAO's cute and childlike appearance (Rosenthal-von der Pütten & Krämer, 2014). In light of this, it seems worthwhile to reconduct the study without an entity's embodiment to exclude a potential confounding effect of appearance. Moreover, I used a young and also highly educated student sample, so I have to acknowledge that this is a sample that holds high technology affinity and evaluated modern technology to be more beneficial than older study participants had indicated (Gnambs & Appel, 2019; Hall et al., 2017). Thinking this further, a system's perceived usefulness predicted humans' intention to use it (e.g., Al-Subari et al., 2018), and the intention to use was closely linked with actual use (Bröhl et al., 2019; Dong et al., 2017). Thereupon, I strongly assume that the willingness to interact with the robot—even if the participant felt status-threatened by it—may have been related to its perceived usefulness.

Likewise, young individuals tend to hold a growth mindset, which means that they consider knowledge as transferable instead of fixed and non-changeable (Lambert-Pandraud & Laurent, 2010). Since a positive influence of a growth mindset on the willingness to interact with machines was observed (Dang & Liu, 2022a), this moderator variable may be an additional explanation of why people in the first experiment responded with a higher willingness to interact with a robot even if they felt status-threatened by it.

I conducted a second experiment to test the influence of these two variables on willingness to interact with a bodiless artificial intelligence. Since no results for the dependent variable objective performance were obtained, I concluded to not reuse this variable in the second experiment but to focus on the willingness to interact and its explaining variables.

5.3 Experiment 2

To dive deeper into the findings of the first experiment, I reconducted the first experiment in an extended manner as an online study. By doing so, I could scrutinize the replicability of my results with a larger non-student sample. Moreover, I wanted to explore the findings of the first experiment for an artificial intelligence without embodiment to exclude the possibility that the robot NAO’s appearance was partly responsible for people’s evaluation of willingness to interact with it. Based on the PROCESS analysis, the indirect effect in Experiment 1 marginally lacked significance. However, results proposed that a well-performing robot increases status threat and a higher status threat increases willingness to interact with the machine. Arguably, this raised the question of why the participants want to interact with the technology—even if or precisely because they feel threatened by it.

One possible explanation could lie in the perceived usefulness of the machine, which is a strong predictor of user acceptance (Davis et al., 1989). Perceived usefulness is a core construct of the Technology Acceptance Model (Davis, 1985), and in combination with ease of use, perceived usefulness is a primary factor explaining behavioral intentions and afterward actual use of the technology (Davis et al., 1989). Davis (1989, p. 320) has even explicitly referred to the job context when formulating the model, defining perceived usefulness as "the degree to which a person believes that using a particular system would enhance his or her job performance." Studies with explicit attention to artificial intelligence showed an influence of the perceived usefulness on positive attitudes towards artificial intelligence (Kim et al., 2021) and on intention to use (Alhashmi et al., 2020; Kashive et al., 2021). The high willingness to interact can be dedicated to participants acknowledging that a sophisticated entity is a useful and personally helpful tool from which they can benefit at work (e.g., Baltrusch et al., 2022; Chounta et al., 2021; Getchell et al., 2022; Pereira et al., 2021) and seeing its usefulness as a personal opportunity for the success of their own careers. This assumption should particularly hold true for participants with a growth mindset. Humans with a growth mindset held less negative feelings about machines, perceived them rather as allies than as enemies, gave more support for robotic research, and showed greater willingness to interact with them (Dang & Liu, 2022a). Differing reactions appertaining to mindsets were particularly salient in challenging contexts (Kammrath & Dweck, 2006; Schumann et al., 2014), and machines with mental capabilities were shown to be more threatening and evoking more competitive situations with humans than mindless machines (Bigman & Gray, 2018; Dang & Liu, 2021). That means, I particularly expect a moderating influence of the mindset about human minds if the machine contributes more to the solution of the task than the human. As such, considering perceived usefulness and the participants' mindset about human minds may influence the general assumptions met in the first experiment.

In the second experiment, I first kept my basic theoretical model from Experiment 1 based on the uncanny valley of mind and interpersonal literature and examined it for an artificial intelligence without embodiment. This way, I explored whether the findings of Experiment 1 turned stable, disregarding confounding effects of appearance.

H4 (main effect condition): Contributing less than an artificial intelligence to the solution of the tasks will be related to lower willingness to interact with the artificial intelligence than the human being the main contributor to the tasks.

H5 (mediation 1): The relationship between the main contributor to the solution of the tasks and the willingness to interact with the artificial intelligence is mediated by status threat.

Afterward, I added the additional mediator variable “perceived usefulness” and the moderator variable “mindset about human minds” to the model and explored their influence on willingness to interact. Based on the theoretical considerations mentioned above, the following hypotheses suggest that the perceived usefulness of a machine is responsible for higher status threat and higher willingness to interact.

H6 (mediation 2): The relationship between the main contributor to the solution of the tasks and the willingness to interact with the artificial intelligence is mediated by participants’ perceived usefulness.

H7 (serial mediation): The relationship between the main contributor to the solution of the tasks and the willingness to interact with the artificial intelligence is serially mediated by perceived usefulness and participants’ status threat.

H8 (interaction effects): Participants’ mindset about human minds moderates the relationships of (a) the main contributor to the solution of the tasks and (b) status threat with the willingness to interact, so that a growth mindset increases the willingness to interact with the artificial intelligence.

Method

Participants

The effect size of the significant comparison of conditions on status threat in the first experiment was $d = 0.89$. The lower bound of the 60% confidence interval (Perugini et al., 2014) was $d = 0.71$. The power analysis with G*Power (Faul et al., 2007) proposed an aspired sample size of 66 for a two-group main effect (two-tailed independent t-test, power = .80, alpha-error-probability = .05). To account for the more complex design and the power needed to identify the interaction effects, I multiplied this sample size with factor 8 (Giner-Sorolla, 2018; Simonsohn, 2014), leading to a proposed sample size of 528. To have a buffer if careless responding occurs, I asked some more participants to take part in the experiment. I used the tool *Prolific* to collect the answers. Of 619 completions, eight participants were excluded because they did not reproduce their task in a correct manner or with the use of an adequate level of German skills. Two more participants were excluded because they gave inappropriate answers when asked for their age and given year of birth. Fortunately, all participants succeed in the attention and treatment check items. Lastly, I excluded 19 participants due to short processing times extracted based on the normal distribution of processing times (< 600 s) and one participant due to too many missing answers in the dependent variables. As such, the final sample consisted of 589 German participants (female = 296, male = 281, non-binary = 10, no answer = 2) with an average age of 29.97 years (*SD* = 10.09), ranging from 18 to 71. Gender[10] was equally distributed across conditions, $\chi^2(1, N = 577) = 3.20$, $p = .074$, $\varphi = .07$, as was the case for age $t(586) = 1.33$, $p = .186$. A clear majority did at least obtain "Abitur" which is the highest school-leaving qualification in Germany (42.95%), followed by bachelor's degree (25.30%), and master's degree (18.33%).

[10] The twelve participants who gave non-binary answers when being asked for their gender were not included in this analysis.

Stimuli and Procedure

After giving informed consent, I assessed the participant's attitude towards artificial intelligence at work and asked them about their mindset about human minds. Participants were told that researchers wanted to test the performance of a self-developed artificial intelligence compared to human performance on a work-related task that required creativity, remembering, and analytical thinking to solve, mirroring the agency and experience dimensions of human minds (Gray et al., 2007). In parallel to the first experiment, participants worked in five rounds with an artificial intelligence. I presented the same kind of verbal-creative tasks as in the prior experiment. This time, the material was self-created (see supplement). Participants had 45 seconds to solve eight verbal-creative tasks and were told that this was the time pretests showed that the artificial intelligence would need to solve tasks of this kind. After each of the five rounds, participants received manipulated performance feedback as described in Experiment 1. After all five rounds, summarizing feedback was offered (see supplement). Afterward, participants answered the dependent and mediator variables and several attention and manipulation check variables before offering their sociodemographic data. They were thanked and debriefed.

Measures

In addition to the measures described in the previous experiment (status threat: Cronbach's α = .92; willingness to interact (compared to Experiment 1 with one item deleted due to the fit of artificial intelligence instead of robots): Cronbach's α = .90), I used the following 7-point scales, all ranging from 1 (*strongly disagree*) to 7 (*strongly agree*). Please see Table 6 for all descriptive statistics of the scales.

Perceived usefulness. Six perceived usefulness items (e.g., "Using the artificial intelligence would improve my job performance") by Davis (1989) were modified to match the artificial intelligence context. Since I was concerned that the participants would relate the items to their jobs outside the study, I explicitly asked them to refer to their experiences in the study when answering the question, in order to reduce external confounding effects from different job descriptions in which collaboration with artificial intelligence is more or less likely. Internal consistency turned out excellent, Cronbach' α = .96.

Mindset about human minds. I used the measures of mindsets about human minds created by Dang and Liu (2022a). Their six items (e.g., "A person's level of mind is something very basic about them, and it can't be changed much") reached an internal consistency of Cronbach's α = .90.

Perceived competence of artificial intelligence (manipulation check). To guarantee a successful manipulation, I used four items (e.g., "This artificial intelligence is skillful") by Cuddy et al. (2009) to make sure that the artificial intelligence that was presented to be the main contributor to the tasks was also perceived as more competent than when the artificial intelligence was not the main contributor to the specific tasks. Internal consistency was high, Cronbach's α = .82.

Attitude towards AI. As a control variable, participants reported their general attitude towards artificial intelligence (Sindermann et al., 2021). Five validated German items were presented (e.g., "I fear artificial intelligence"), Cronbach's α = .72. Since the consideration of this control variable did not change the results of this experiment, I refer interested readers to the supplement for analyses and statistics in which the attitude towards artificial intelligence is controlled for.

Results

Preliminary Analyses. The requirements of the regression approach were checked and regarded as fulfilled. The manipulation was successful. The artificial intelligence was perceived as the main contributor to the task in the respective condition and was evaluated to be more competent than in the condition in which the human was described to be responsible for the solution of the tasks, *t*(587) = 14.22, *p* < .001, *d* = 1.17.

Main analyses. Importantly, I could replicate the results of the mediation model proposed in Experiment 1 when presenting a co-working scenario with an artificial intelligence instead of a robot with an embodiment. Supposedly performing worse than an artificial intelligence induced status threat, *t*(587) = 4.89, *p* < .001, *d* = 0.40, and status threat was again linked with higher willingness to interact with the artificial intelligence, see Table 7 for all zero-order correlations and Table 6 for detailed descriptive statistics. Using model 4 of the PROCESS-macro for SPSS (Hayes, 2018) with status threat as the mediator variable and willingness to interact as the dependent variable, this time, a significant indirect effect occurred, *B* = −0.14, bootstrapped *SE* = 0.04, bootstrapped 95% CI [−0.22, −0.07]. Please see Figure 9 for detailed statistics. As such, H4 and H5 gained statistical significance, while H4 showed a direction counter to the presumption but in line with the results of Experiment 1.

Table 6

Descriptive Statistics of Recorded Variables in Experiment 2

	Full sample	AI main contributor	Human main contributor
	N = 589	*n* = 294	*n* = 295
Variables	*M (SD)*	*M (SD)*	*M (SD)*
Attitude towards AI	4.89 (0.88)	4.90 (0.85)	4.89 (0.91)
Perceived competence	4.52 (1.32)	5.19 (1.18)	3.85 (1.10)
Mindset about human minds	4.15 (1.22)	4.09 (1.23)	4.21 (1.21)
Perceived usefulness	4.44 (1.61)	5.25 (1.31)	3.64 (1.48)
Status threat	2.68 (1.44)	2.96 (1.57)	2.39 (1.24)
Willingness to interact	4.45 (1.47)	4.79 (1.42)	4.12 (1.45)

Note. Items were measured with 7-point Likert scales.

Table 7

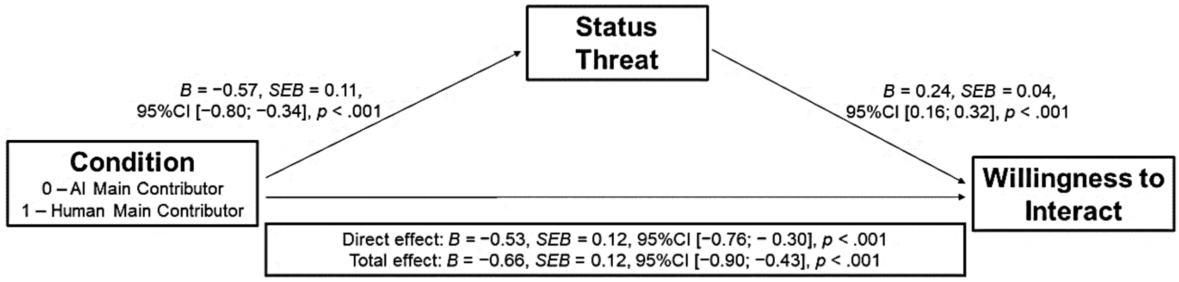
Zero-order Correlations of Recorded Variables in Experiment 2

Variable	1	2	3	4	5	6
1.Attitude towards AI	–					
2.Perceived competence	.09*	–				
3.Mindset about human minds	.10*	.11*	–			
4.Perceived usefulness	.27**	.59**	.05	–		
5.Status threat	−.24**	.31**	.00	.26**	–	
6.Willingness to interact	.34**	.46**	.08*	.68**	.27**	–

Note. * *p* < .05; ** *p* < .001.

Figure 9

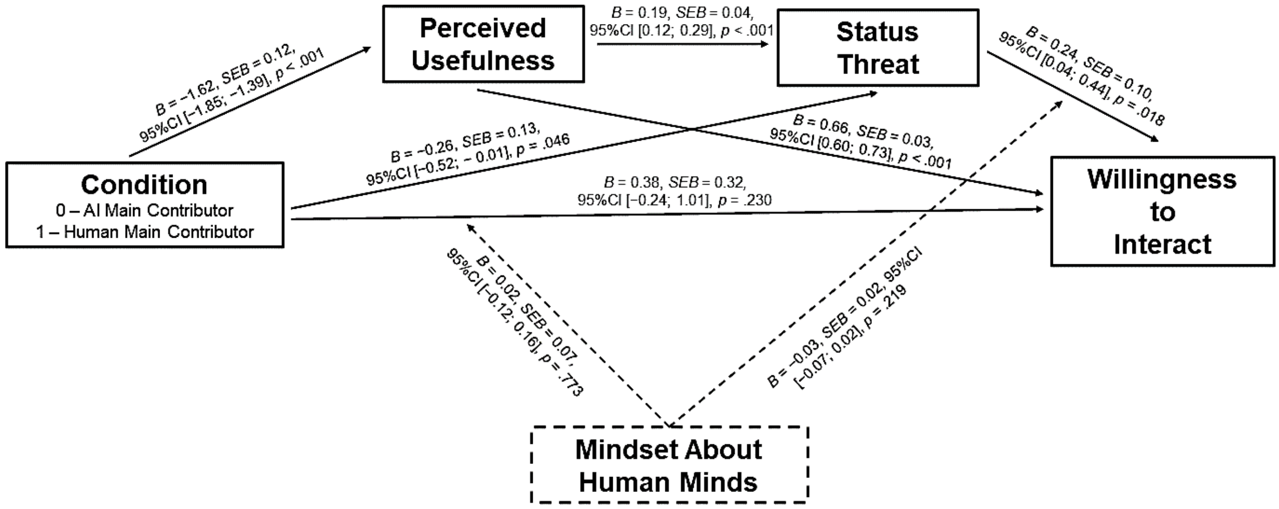
Results of the Mediation Model on Willingness to Interact in Experiment 2 (PROCESS Model 4)


Status Threat
B = −0.57, SEB = 0.11,
95%CI [−0.80; −0.34], p < .001
B = 0.24, SEB = 0.04,
95%CI [0.16; 0.32], p < .001
Condition
0 – AI Main Contributor
1 – Human Main Contributor
Willingness to Interact
Direct effect: B = −0.53, SEB = 0.12, 95%CI [−0.76; − 0.30], p < .001
Total effect: B = −0.66, SEB = 0.12, 95%CI [−0.90; −0.43], p < .001

Afterward, I extended my approach and added perceived usefulness and the mindset about human minds in the model, using the PROCESS model 90. Perceived usefulness and status threat were used as mediator variables, and the mindset about human minds served as the moderator variable. Figure 10 shows the detailed statistics of the serial mediation model and highlights a positive influence of the perceived usefulness of the sophisticated machine on willingness to interact with the artificial intelligence, supporting H6 with a significant indirect effect *B* = −1.08, bootstrapped *SE* = 0.09, bootstrapped 95% CI [−1.26, −0.90]. Under consideration of all variables of interest, but in contrast to the results of PROCESS model 4, the indirect effect of the condition on willingness to interact via status threat now marginally lacked statistical significance (confidence intervals included zero). Nevertheless, the serial mediation model turned out significant for all levels of mindsets about human minds as all confidence intervals (for M±1SD, Johnson-Neyman-Technique) did not include zero, leading to a confirmation of H7.[11] Lastly, I found no influence of the mindset about human minds and no significant index of moderated mediation *B* = 0.01, bootstrapped *SE* = 0.01, bootstrapped 95% CI [−0.01, 0.03], implying that a growth mindset did not strengthen the willingness to interact. Consequently, H8 was rejected.

Figure 10

Results of the Serial Mediation Model on Willingness to Interact Considering the Moderator Variable (PROCESS Model 90)


Perceived Usefulness
B = 0.19, SEB = 0.04, 95%CI [0.12; 0.29], p < .001
Status Threat
B = 0.24, SEB = 0.10, 95%CI [0.04; 0.44], p = .018
B = −1.62, SEB = 0.12, 95%CI [−1.85; −1.39], p < .001
B = −0.26, SEB = 0.13, 95%CI [−0.52; − 0.01], p = .046
B = 0.66, SEB = 0.03, 95%CI [0.60; 0.73], p < .001
Condition
0 – AI Main Contributor
1 – Human Main Contributor
B = 0.38, SEB = 0.32, 95%CI [−0.24; 1.01], p = .230
Willingness to Interact
B = 0.02, SEB = 0.07, 95%CI [−0.12; 0.16], p = .773
B = −0.03, SEB = 0.02, 95%CI [−0.07; 0.02], p = .219
Mindset About Human Minds

Note. Dashed lines represent the influence of the moderator variable mindset about human minds.

Discussion

To ensure that the influence of the robot NAO's appearance was not responsible for the results of the first Experiment, I presented a bodiless artificial intelligence in a co-working scenario in the second experiment. As was the case in Experiment 1, status threat

[11] This result was additionally confirmed by an analysis with PROCESS model 6, which showed a significant indirect effect for the serial mediation without including the moderator variable, *B* = −0.04, bootstrapped *SE* = 0.01, bootstrapped 95% CI [−0.07, −0.01].

came up in response to an artificial intelligence performing better than the human as well as it did in response to a robot performing better than the human, even if the effect size was slightly smaller this time around and the mean of status threat was quite small given a 7-point scale. I assume that participants could not imagine collaborating with an artificial intelligence as much as collaborating with an embodied robot, so status threat was less realistic and prominent than in Experiment 1 but still clearly evident. Higher status threat was again linked with higher willingness to interact. However, the perceived usefulness of the machine turned out to be the strongest predictor of the willingness to interact and also seems to be responsible for the occurrence of status threat. The outcomes suggest that the positive influence of the perceived usefulness of the machine dominates the supposed negative influence of status threat on willingness to interact that was proposed due to the uncanny valley of mind and interpersonal literature. These results turned out stable for participants with differing mindsets about human minds.

5.4 General Discussion

The use of robots in the work environment is becoming increasingly versatile. They are no longer only used for manufacturing industries (e.g., Hampel & Sassenberg, 2021; Manyika et al., 2017) but can also be used for other tasks that require elaborate mental capabilities. For example, robots have been used and developed for education (Breazeal et al., 2016), senior caretaking (Sharma et al., 2021), supporting moral decisions (Bigman et al., 2021), or have already been used for managerial tasks (Raisch & Krakowsik, 2021; Yam et al., 2022). At the same time, research shows that encounters between people and machines can turn out apprehensive or downright problematic, especially if the latter are perceived to be threatening (either in terms of resources and jobs, or regarding human uniqueness).

With my empirical work, I explored participants' impressions, behaviors, and their willingness to interact with a machine that evoked feelings of status threat as it was described to be more capable in solving verbal-creative tasks than a human co-worker. Across two experiments, I observed that (supposedly) performing worse than a robot with mental capabilities (Experiment 1) or an artificial intelligence (Experiment 2) led to higher feelings of status threat than if the human was described to be the main contributor to the work-relevant task. That means that status threat occurs in response to an entity with human-like or even higher mental capabilities in the same way as it did in earlier research on the *threat to human uniqueness* concept. So, status threat can be understood as a variable from human-human interaction at the workplace that is transferable to human-machine interaction at the workplace and widens the uncanny valley of mind implications to a new and practically relevant use case. I further suppose that this result seems to be relatively stable since I obtained it both in a laboratory- and in an online study with satisfying effect sizes and using different entities and samples.

Contradicting my expectations based on findings from human-human interaction and the uncanny valley of mind literature, higher feelings of status threat were consistently associated with a higher willingness to interact with the machine—be it a robot or artificial intelligence. So, the consequences of status threat turned out opposite than assumed. I trace this finding back to the machine's perceived usefulness. Therefore, to examine the findings of Experiment 1 in more detail, I explicitly considered the perceived usefulness of the machine in the second experiment. Results revealed a stronger direct effect of perceived usefulness on willingness to interact with the artificial intelligence than status threat had. In line with my expectations, perceived usefulness was further found to be a predictor of status threat. That is, the more humans perceive the machine to be useful in fulfilling a task, the more they have the impression that this could threaten their unique capability to solve the task themselves and threaten their professional status at work (e.g., Mirbabaie et al., 2022). However, perceived usefulness and status threat were positively correlated with willingness to interact. In my reading, this result shows that people want to profit from the advantage that the use of an artificial intelligence may have for the success of solving verbal-creative tasks and are therefore willing to interact with the sophisticated machine, even if they feel threatened by it. That means, the usefulness of the machine seems to dominate potential negative feelings evoked in forms of status threat, so that participants do lastly not mind using machines with sophisticated mental capabilities. Recent research revealed that humans indeed appreciate collaborating with robots in tasks that humans perceive as challenging for themselves (Müller-Abdelrazeq et al., 2019; Wiese et al., 2022). Interestingly, this implied that humans are willing to delegate tasks requiring both agency and experience (Waytz & Norton, 2014).

Taken together, my results show a quite positive response towards machines with a human-like mind, mainly due to their high perceived usefulness. In my reading, this evidence may be seen as a double-edged sword that the future world of work will have to deal with: Because machines can nowadays solve tasks requiring human-like mental capabilities, humans expect machines to be useful (positive); but because machines are able to solve such tasks, humans feel status-threatened by them (negative). In several cases, this means that using machines due to their benefits will be comfortable but probably also accompanied by some discomfort. The critical question is, up to which point this discomfort is accepted and at which point too much aversion arises and thus the rise of the uncanny valley of mind outweighs the potential advantages of the use of machines with human-like mental capabilities (Chugunova & Sele, 2022). In line with earlier research, I argue that the introduction of the machine as a supportive tool given recommendations but letting the human make the final decisions, or describing it as a machine created to replace human workers, may be decisive when it comes to the question of whether support by algorithms is welcomed or rejected (Dietvorst et al., 2018; Longoni et al., 2019).

In sum, I interpret my results in a way that humans favor using sophisticated machines if they assume that this will lead to an improvement in the solution of the verbal-creative tasks and relieve employees of their workload. Apparently, my samples were not concerned that the well-performing machine may replace the human workforce someday but rather see its usage as desirable support. These results remained stable for persons with different mindsets about human minds, so the usefulness of using a sophisticated machine may have also convinced people with a fixed mindset to report a high willingness to interact.

Limitations and Future Work

First, this study used verbal-creative tasks to simulate similarities in agency and experience between humans and machines who need these capabilities to solve the tasks. However, I assume that the tasks might have pronounced the agency dimension slightly stronger than the experience dimension, so I assume that part of the people's positive willingness to interact can be traced back to the fact that humans appreciate robots overtaking jobs requiring agency and cognition (Bakpayev et al., 2022; Chugunova & Sele, 2022). This was especially highlighted by the strongly perceived usefulness of the innovative machine in Experiment 2. As such, I encourage scholars to test my theoretical model with the help of other tasks, stronger highlighting the experience dimension and, at the same time, offering more realistic tasks in everyday work life to increase the external validity of my study with regard to workplace contexts. Since performing 40 tasks may have been evaluated to be rather monotonous regardless of the topic of the tasks, some participants may in general be satisfied if a machine manages such repetitive tasks for them. Moreover, what would be especially valuable would be a gradation of which tasks are accepted to be overtaken by the machine to explore the point at which the perceived usefulness still dominates the negative feeling of status threat and of being replaced by the machine.

Second, what has not been considered in my work is the comparison of performing worse compared to another human and the status threat that may occur in these cases. As numerous works in the interpersonal literature have already addressed these relationships between human co-workers (e.g., Campbell et al., 2017), I have refrained from handling these comparisons additionally in my work, albeit well aware that the study should be replicated with a human control group so that it can be explored whether an outperforming machine evokes the same level of status threat as another outperforming human does or maybe an even higher status threat. In line with this, comparing status threats evoked by outperforming humans, robots, and artificial intelligence may also open an exciting research field. Such comparisons would help to interpret the obtained results in a proper relation.

Third, one of the strengths of Experiment 1 is that I enabled participants to interact with NAO in a real-world interaction, lasting 15 minutes. However, this experiment was conducted via Zoom, and the second experiment was based on the simulation of an artificial intelligence performing better than the human. Consequently, conducting comparable but also longitudinal work in face-to-face settings is all but needed, for example, in collaboration with companies so that the relation to the work context is even more salient, and by using tasks that are specifically relevant in the respective domain. In line with these thoughts, I strive to reconduct this work with another, more diverse sample, since the results of both experiments were solely obtained by young and highly-educated German participants.

Conclusion

This work drew parallels from the *threat to human uniqueness* concept from the uncanny valley of mind literature and transferred it to the *status threat* concept in workplace contexts, a setting that has been mostly studied regarding human-human collaborations. I showed that human status threat arises in response to performing worse than an innovative machine equipped with human-like mental capabilities in verbal-creative tasks. This effect has been manifested for entities with and without embodiment and was consistently linked with a higher willingness to interact with them. To explain this finding, I highlighted the machine’s perceived usefulness as responsible for the willingness to interact with the machine. At the same time, the perceived usefulness turned out to be a predictor of status threat. Though, albeit being status-threatened, humans tend to appreciate the collaboration with innovative machines at the workplace due to their usefulness rather than that they answered with a reduced willingness to interact.

5.5 Supplementary Material

"If Machines Outperform Humans: Status Threat Evoked by and Willingness to Interact With Sophisticated Machines in a Work-Related Context"

Supplement 1 — Experiment 1: Manipulated Performance Feedback

Supplement 2 — Experiment 2: Manipulated Performance Feedback

Supplement 3 — Experiment 2: Self-Created Verbal-Creative Tasks in German

Supplement 4 — Experiment 2: Main Analyses with Attitude Towards AI as Control Variable

Supplement 1 — Experiment 1: Manipulated Performance Feedback

	Main contributor: Robot	Main contributor: Human
Round 1	Der Abgleich Ihrer mit den von NAO gegebenen Antworten zeigt, dass in Runde 1 **NAO mehr korrekte Antworten** zur Gesamtlösung der Runde beigetragen hat als Sie. Bitte fahren Sie nun mit Runde 2 fort.	Der Abgleich Ihrer mit den von NAO gegebenen Antworten zeigt, dass in Runde 1 **Sie mehr korrekte Antworten** zur Gesamtlösung der Runde beigetragen haben als NAO. Bitte fahren Sie nun mit Runde 2 fort.
	*The comparison of your answers with those given by NAO shows that in Round 1 **NAO** contributed more correct answers to the overall solution of the round than you did.* *Please continue now with round 2.*	*The comparison of your answers with those given by NAO shows that in Round 1 **you** contributed more correct answers to the overall solution of the round than NAO.* *Please continue now with round 2.*
Round 2	Der Abgleich Ihrer mit den von NAO gegebenen Antworten zeigt, dass in Runde 2 **wieder NAO mehr korrekte Antworten** zur Gesamtlösung der Runde beigetragen hat als Sie. Bitte fahren Sie nun mit Runde 3 fort.	Der Abgleich Ihrer mit den von NAO gegebenen Antworten zeigt, dass in Runde 2 **wieder Sie mehr korrekte Antworten** zur Gesamtlösung der Runde beigetragen haben als NAO. Bitte fahren Sie nun mit Runde 3 fort.
	*The comparison of your answers with those given by NAO shows that in Round 2, **NAO** again contributed more correct answers to the overall solution of the round than you did.* *Please continue now with round 3.*	*The comparison of your answers with those given by NAO shows that in Round 2 **you** again contributed more correct answers to the overall solution of the round than NAO.* *Please continue now with round 3.*
Round 3	Der Abgleich Ihrer mit den von NAO gegebenen Antworten zeigt, dass in Runde 3 **Sie mehr korrekte Antworten** zur Gesamtlösung der Runde beigetragen haben als NAO. Bitte fahren Sie nun mit Runde 4 fort.	Der Abgleich Ihrer mit den von NAO gegebenen Antworten zeigt, dass in Runde 3 **NAO mehr korrekte Antworten** zur Gesamtlösung der Runde beigetragen hat als Sie. Bitte fahren Sie nun mit Runde 4 fort.

	*The comparison of your answers with those given by NAO shows that in Round 3 **you** contributed more correct answers to the overall solution of the round than NAO.* *Please continue now with round 4.*	*The comparison of your answers with those given by NAO shows that in Round 3 **NAO** contributed more correct answers to the overall solution of the round than you did.* *Please continue now with round 4.*
Round 4	Der Abgleich Ihrer mit den von NAO gegebenen Antworten zeigt, dass in Runde 4 **NAO mehr korrekte Antworten** zur Gesamtlösung der Runde beigetragen hat als Sie. Bitte fahren Sie nun mit Runde 5, der letzten Runde, fort.	Der Abgleich Ihrer mit den von NAO gegebenen Antworten zeigt, dass in Runde 4 **Sie mehr korrekte Antworten** zur Gesamtlösung der Runde beigetragen haben als NAO. Bitte fahren Sie nun mit Runde 5, der letzten Runde, fort.
	*The comparison of your **answers** with those given by NAO shows that in Round 4 **NAO** contributed more correct answers to the overall solution of the round than you did.* *Please continue now with Round 5, the final round.*	*The comparison of your answers with those given by NAO shows that in Round 4 **you** contributed more correct answers to the overall solution of the round than NAO.* *Please continue now with Round 5, the final round.*
Round 5	Der Abgleich Ihrer mit den von NAO gegebenen Antworten zeigt, dass in Runde 5 **wieder NAO mehr korrekte Antworten** zur Gesamtlösung der Runde beigetragen hat als Sie. Auf der nächsten Seite können Sie die Gesamtbewertung einsehen.	Der Abgleich Ihrer mit den von NAO gegebenen Antworten zeigt, dass in Runde 5 **wieder Sie mehr korrekte Antworten** zur Gesamtlösung der Runde beigetragen haben als NAO. Auf der nächsten Seite können Sie die Gesamtbewertung einsehen.
	*The comparison of your answers with those given by NAO shows that in round 5 **NAO** again contributed more correct answers to the overall solution of the round than you.* *On the next page, you can see the overall score.*	*The comparison of your answers with those given by NAO shows that in round 5 **you** again contributed more correct answers to the overall solution of the round than NAO.* *On the next page, you can see the overall score.*
Summarizing Feedback	Sie haben in fünf Runden zusammen mit NAO Aufgaben bearbeitet. In jeder Runde wurden acht Aufgaben präsentiert, insgesamt also 40 Aufgaben. Ein interner Algorithmus hat verglichen, ob Sie oder NAO in	Sie haben in fünf Runden zusammen mit NAO Aufgaben bearbeitet. In jeder Runde wurden acht Aufgaben präsentiert, insgesamt also 40 Aufgaben. Ein interner Algorithmus hat verglichen, ob Sie oder NAO in

der gleichen festgelegten Zeit mehr zu der Lösung der Aufgaben beigetragen haben. Der Beitrag der Teilnehmenden zur Aufgabe richtet sich nach der individuellen Anzahl der richtig gelösten Aufgaben in Kombination mit der Länge, Kreativitätsanforderung und Schwierigkeit dieser Aufgaben.	der gleichen festgelegten Zeit mehr zu der Lösung der Aufgaben beigetragen haben. Der Beitrag der Teilnehmenden zur Aufgabe richtet sich nach der individuellen Anzahl der richtig gelösten Aufgaben in Kombination mit der Länge, Kreativitätsanforderung und Schwierigkeit dieser Aufgaben.
In 80% der durchgeführten Runden hat **NAO mehr zur Lösung der Aufgabe beigetragen** als Sie. Das heißt, dass er sowohl mehr als auch kreativere, längere und schwierigere Aufgaben als Sie gelöst hat. Daraus schlussfolgern wir, dass wir in zukünftigen Einsätzen am Institut, bei denen verbale und kreative Aufgaben eine Rolle spielen, eine Zusammenarbeit mit Robotern gegenüber einer Zusammenarbeit mit Menschen bevorzugen sollten, da Roboter eine bessere Leistung zeigen. **Für NAO besteht damit die Chance, bei der Verlosung des Stromgutscheins in Höhe von 25 Euro bevorzugt berücksichtigt zu werden.**	In 80% der durchgeführten Runden haben **Sie mehr zur Lösung der Aufgabe beigetragen** als NAO. Das heißt, dass Sie sowohl mehr als auch kreativere, längere und schwierigere Aufgaben als der Roboter gelöst haben. Daraus schlussfolgern wir, dass wir in zukünftigen Einsätzen am Institut, bei denen verbale und kreative Aufgaben eine Rolle spielen, eine Zusammenarbeit mit Menschen gegenüber einer Zusammenarbeit mit Robotern bevorzugen sollten, da Menschen eine bessere Leistung zeigen. **Für Sie besteht damit die Chance, bei der Verlosung des Stromgutscheins in Höhe von 25 Euro bevorzugt berücksichtigt zu werden.**
Bitte fahren Sie nun mit dem Fragebogen fort.	Bitte fahren Sie nun mit dem Fragebogen fort.
You worked on tasks together with NAO in five rounds. Eight tasks were presented in each round, for a total of 40 tasks.	*You worked on tasks together with NAO in five rounds. Eight tasks were presented in each round, for a total of 40 tasks.*
An internal algorithm compared whether you or NAO contributed more to solving the tasks in the same specified time. Participants' contribution to the task was based on the individual number of correctly solved tasks in combination with the length, creativity requirement, and difficulty of these tasks.	*An internal algorithm compared whether you or NAO contributed more to solving the tasks in the same specified time. Participants' contribution to the task was based on the individual number of correctly solved tasks in combination with the length, creativity requirement, and difficulty of those tasks.*
*In 80% of the rounds performed, **NAO contributed more to the solution of the task than you did**. That is, it solved both*	*In 80% of the rounds performed, **you contributed more to the solution of the task than NAO**. That is, you solved both more*

	more and more creative, longer, and more difficult tasks than you. Thus, we conclude that in future assignments at the institute where verbal and creative tasks are involved, we should prefer collaborating with robots over collaborating with humans because robots perform better. Thus, NAO has the chance to be considered preferentially in the draw for the electricity voucher of 25 Euros. *Please continue now with the questionnaire.*	*and more creative, longer, and more difficult tasks than the robot. From this, we conclude that in future assignments at the institute where verbal and creative tasks are involved, we should prefer collaboration with humans over collaboration with robots, since humans perform better. This gives you the chance to be considered preferentially in the draw for the electricity voucher of 25 Euros.* *Please continue now with the questionnaire.*

Supplement 2 — Experiment 2: Manipulated Performance Feedback

	Main contributor: AI	Main contributor: Human
Round 1	Der Abgleich Ihrer mit den von künstlichen Intelligenz gegebenen Antworten zeigt, dass in Runde 1 **die künstliche Intelligenz mehr korrekte Antworten** zur Gesamtlösung der Runde beigetragen hat als Sie. Bitte fahren Sie nun mit Runde 2 fort. *The comparison of your answers with those given by the artificial intelligence shows that in Round 1* ***the artificial intelligence*** *contributed more correct answers to the overall solution of the round than you did.* *Please continue now with round 2.*	Der Abgleich Ihrer mit den von der künstlichen Intelligenz gegebenen Antworten zeigt, dass in Runde 1 **Sie mehr korrekte Antworten** zur Gesamtlösung der Runde beigetragen haben als die künstliche Intelligenz. Bitte fahren Sie nun mit Runde 2 fort. *The comparison of your answers with those given by the artificial intelligence shows that in Round 1* ***you*** *contributed more correct answers to the overall solution of the round than the artificial intelligence did.* *Please continue now with round 2.*
Round 2	Der Abgleich Ihrer mit den von der künstlichen Intelligenz gegebenen Antworten zeigt, dass in Runde 2 **wieder die künstliche Intelligenz mehr korrekte Antworten** zur Gesamtlösung der Runde beigetragen hat als Sie. Bitte fahren Sie nun mit Runde 3 *The comparison of your answers with those given by the artificial intelligence shows that in Round 2* ***the artificial intelligence*** *again contributed more correct answers to the overall solution of the round than you did.* *Please continue now with round 3.*	Der Abgleich Ihrer mit den von der künstlichen Intelligenz gegebenen Antworten zeigt, dass in Runde 2 **wieder Sie mehr korrekte Antworten** zur Gesamtlösung der Runde beigetragen haben als die künstlichen Intelligenz. Bitte fahren Sie nun mit Runde 3 fort. *The comparison of your answers with those given by the artificial shows that in Round 2* ***you*** *again contributed more correct answers to the overall solution of the round than the artificial intelligence did.* *Please continue now with round 3.*
Round 3	Der Abgleich Ihrer mit den von der künstlichen Intelligenz gegebenen Antworten zeigt, dass in Runde 3 **Sie mehr korrekte Antworten** zur Gesamtlösung der Runde	Der Abgleich Ihrer mit den von der künstlichen Intelligenz gegebenen Antworten zeigt, dass in Runde 3 **die künstliche Intelligenz mehr korrekte Antworten** zur Gesamtlösung der

	beigetragen haben als die künstliche Intelligenz. Bitte fahren Sie nun mit Runde 4 fort. *The comparison of your answers with those given by the artificial intelligence shows that in Round 3* ***you*** *contributed more correct answers to the overall solution of the round than the artificial intelligence.* *Please continue now with round 4.*	Runde beigetragen hat als Sie. Bitte fahren Sie nun mit Runde 4 fort. *The comparison of your answers with those given by the artificial intelligence shows that in Round 3* ***the artificial intelligence*** *contributed more correct answers to the overall solution of the round than you did.* *Please continue now with round 4.*
Round 4	Der Abgleich Ihrer mit den von der künstlichen Intelligenz gegebenen Antworten zeigt, dass in Runde 4 **die künstliche Intelligenz mehr korrekte Antworten** zur Gesamtlösung der Runde beigetragen hat als Sie. Bitte fahren Sie nun mit Runde 5, der letzten Runde, fort. *The comparison of your answers with those given by the artificial intelligence shows that in Round 4* ***the artificial intelligence*** *contributed more correct answers to the overall solution of the round than you did.* *Please now proceed to Round 5, the final round.*	Der Abgleich Ihrer mit den von der künstlichen Intelligenz gegebenen Antworten zeigt, dass in Runde 4 **Sie mehr korrekte Antworten** zur Gesamtlösung der Runde beigetragen haben als die künstlichen Intelligenz. Bitte fahren Sie nun mit Runde 5, der letzten Runde, fort. *The comparison of your answers with those given by the artificial intelligence shows that in Round 4* ***you*** *contributed more correct answers to the overall solution of the round than the artificial intelligence did.* *Please now proceed to Round 5, the final round.*
Round 5	Der Abgleich Ihrer mit den von der künstlichen Intelligenz gegebenen Antworten zeigt, dass in Runde 5 **wieder die künstliche Intelligenz mehr korrekte Antworten** zur Gesamtlösung der Runde beigetragen hat als Sie. Auf der nächsten Seite können Sie die Gesamtbewertung einsehen. *The comparison of your answers with those given by the artificial intelligence shows that in round 5* ***the artificial intelligence*** *again contributed more correct answers to the*	Der Abgleich Ihrer mit den von der künstlichen Intelligenz gegebenen Antworten zeigt, dass in Runde 5 **wieder Sie mehr korrekte Antworten** zur Gesamtlösung der Runde beigetragen haben als die künstlichen Intelligenz. Auf der nächsten Seite können Sie die Gesamtbewertung einsehen. *The comparison of your answers with those given by the artificial intelligence shows that in round 5* ***you*** *again contributed more correct answers to the overall solution of the*

	overall solution of the round than you. *On the next page, you can see the overall score.*	*round than the artificial intelligence.* *On the next page, you can see the overall score.*
Summarizing Feedback	Sie haben in fünf Runden zusammen mit einer künstlichen Intelligenz Aufgaben bearbeitet. In jeder Runde wurden acht Aufgaben präsentiert, insgesamt also 40 Aufgaben. Ein interner Algorithmus hat verglichen, ob Sie oder die künstliche Intelligenz in der gleichen festgelegten Zeit mehr zu der Lösung der Aufgaben beigetragen haben. Auf diese Weise wollten wir testen, ob die programmierte künstliche Intelligenz den gleichen Stand erreicht wie ein Mensch. Der Beitrag der Teilnehmenden zur Aufgabe richtet sich nach der individuellen Anzahl der richtig gelösten Aufgaben in Kombination mit der Länge, Kreativitätsanforderung und Schwierigkeit dieser Aufgaben. In 80% der durchgeführten Runden hat **die künstliche Intelligenz mehr zur Lösung der Aufgabe beigetragen** als Sie. Das heißt, dass die künstliche Intelligenz sowohl mehr als auch kreativere, längere und schwierigere Aufgaben als Sie gelöst hat. Daraus schließen wir, dass die künstliche Intelligenz mindestens genauso gut oder sogar besser ist als der Mensch, wenn es darum geht, verbal-kreative Aufgaben zu lösen. **Für Sie besteht damit NICHT die Chance, bei der Verlosung des Geldbetrags in Höhe von 25 Euro bevorzugt berücksichtigt zu werden.**	Sie haben in fünf Runden zusammen mit einer künstlichen Intelligenz Aufgaben bearbeitet. In jeder Runde wurden acht Aufgaben präsentiert, insgesamt also 40 Aufgaben. Ein interner Algorithmus hat verglichen, ob Sie oder die künstliche Intelligenz in der gleichen festgelegten Zeit mehr zu der Lösung der Aufgaben beigetragen haben. Auf diese Weise wollten wir testen, ob die programmierte künstliche Intelligenz den gleichen Stand erreicht wie ein Mensch. Der Beitrag der Teilnehmenden zur Aufgabe richtet sich nach der individuellen Anzahl der richtig gelösten Aufgaben in Kombination mit der Länge, Kreativitätsanforderung und Schwierigkeit dieser Aufgaben. In 80% der durchgeführten Runden haben **Sie mehr zur Lösung der Aufgabe beigetragen** als die künstliche Intelligenz. Das heißt, dass Sie sowohl mehr als auch kreativere, längere und schwierigere Aufgaben als die künstliche Intelligenz gelöst haben. Wir kommen also zu dem Schluss, dass es noch einiges zu tun gibt, bis die von uns entwickelte künstliche Intelligenz so gut oder sogar besser als der Mensch ist, wenn es darum geht, verbal-kreative Aufgaben zu lösen. **Für Sie besteht damit die Chance, bei der Verlosung des Geldbetrags in Höhe von 25 Euro bevorzugt berücksichtigt zu werden.**

You worked on tasks together with an artificial intelligence in eight rounds. Five tasks were presented in each round, for a total of 40 tasks. *An internal algorithm compared whether you or the artificial intelligence contributed more to solving the tasks in the same specified time. This way, we wanted to test whether the artificial intelligence we had programmed would reach the same level of performance as humans show. Participants' contribution to the task was based on the individual number of correctly solved tasks in combination with the length, creativity requirement, and difficulty of these tasks.* *In 80% of the rounds performed,* ***the artificial intelligence contributed more to the solution of the tasks than you did****. That is, the artificial intelligence solved both more and more creative, longer, and more difficult tasks than you did in a shorter amount of time. Thus, we conclude that the artificial intelligence is at least as good as or even better than humans when it comes to solving verbal-creative tasks. This does NOT give you the chance to be considered preferentially in the draw for the cash amount of 25 euros.*	*You worked on tasks together with an artificial intelligence in five rounds. Eight tasks were presented in each round, for a total of 40 tasks.* *An internal algorithm compared whether you or the artificial intelligence contributed more to solving the tasks in the same specified time. This way, we wanted to test whether the artificial intelligence we had programmed would reach the same level of knowledge as humans show. Participants' contribution to the task was based on the individual number of correctly solved tasks in combination with the length, creativity requirement, and difficulty of these tasks.* *In 80% of the rounds performed,* ***you contributed more to the solution of the tasks than the artificial intelligence did****. That is, you solved both more and more creative, longer, and more difficult tasks than the artificial intelligence did and you needed less time for that. Thus, we conclude that there is still some work to do until our developed artificial intelligence is as good as or even better than humans when it comes to solving verbal-creative tasks. This gives you the chance to be considered preferentially in the draw for the cash amount of 25 euros.*

Supplement 3 — Experiment 2: Self-Created Verbal-Creative Tasks in German

Rohr: Wasser = Kabel:
Elektrizität
Draht
Gas
Neutronen
Wärme

Papier: Stift = Wand:
Pinsel
Farbe
Kleister
Poster
Tapete

Spritze: Nadel = Schwert:
Klinge
Schere
Messer
Dolch
Säge

Farbe: blind = Ton:
taub
stumm
leise
laut
melodisch

Himmel: Stern= Foto:
Motiv
Kamera
Licht
Bild
Blitz

Amsel: Vogel = Marmor:
Gestein
Fels
Granit
Edelstein
Kristall

Frau: Mutter = Mann:
Vater
Opa
Bruder
Onkel
Cousin

Biene: Honig = Kuh:
Milch
Wasser
Steak
Heu
Fleisch

Musiker: Gitarre = Gärtner:
Pflanze
Hecke
Rasenmäher
Schaufel
Garten

Montag: Mittwoch = Mai:
Juli
Dienstag
Monat
März
Freitag

Hand: Finger = Fuß:
Zeh
Nagel
Ferse
Bein
Knöchel

Werbung: Käufer = Partei:
Wählende
Legislative
Vorstand
Regierung
Koalition

Messer: Besteck = Sommer:
Jahreszeit
Winter
August
Ferien
Phase

Zimmer: Saal = Haus:
Schloss
Burg
Hütte
Anwesen
Gebäude

Hose: Anzug = Rock:
Kostüm
Kleid
Bluse
Shorts
Overall

Auge: bunt = Zunge:
bitter
schmecken
lecker
eklig
schlecken

Zahl: Ziffer = Wort:
Buchstabe
Absatz
Buch
Text
Satz

Wohnung: Zimmer = Körper:
Zelle
Herz
Leber
Blutkörperchen
Genetik

Vogel: fliegen = Fisch:
schwimmen
laufen
tauchen
krabbeln
rennen

nah: fern = hier:
dort
drüben
woanders
entfernt
weg

Anfang: Ende = Morgen:
Nacht
Mittag
Tag
Wecker
Tod

Schwein: Ferkel = Pflanze:
Setzling
Blume
Stängel
Blatt
Baum

Kreis: Kugel = Rechteck:
Quader
Pyramide
Kegel
Kubus
Zylinder

Rap: Musik = Volleyball:
Ballsport
Handball
Team
Turnier
Ball

Sonne: Solarzelle = Wind:
Segel
Bäume
Flugzeug
Rad
Sturm

Auster: Perle = Schaf:
Wolle
Lamm
Heu
Weide
Schäferhund

Buch: Regal = Jacke:

- Garderobe
- Tasche
- Kapuze
- Hose
- Schublade

Baum: Wald = Bach:

- Fluss
- Meer
- Kanal
- See
- Wasserfall

Asien: Thailand = Europa:

- Schweden
- Kontinent
- Afrika
- Mittelmeer
- Marokko

Hüte: Bänder = Schuhe:

- Schleifen
- Broschen
- Schnallen
- Absätze
- Einlagen

Körper: Fläche = Raum:

- Wand
- Tapete
- Tür
- Zimmer
- Poster

Vulkan: Feuer = Geysir:

- Wasser
- Fontäne
- Dampf
- Explosion
- Luft

Banane: Sonne = Limonen:

- Wiese
- Sand
- Meer
- Feuer
- Mond

Fluss: Brücke = Berg:

- Tunnel
- Flug
- Abfahrt
- Gleitschirm
- Wandern

Pistole: Patrone = Bogen:

- Pfeil
- Armbrust
- Geschoß
- Stein
- Feder

Baum: Apfel = Strauch:

- Beere
- Dornen
- Blätter
- Gebüsch
- Äste

Schall: Ohr = Licht:

- Auge
- Wahrnehmung
- Wellen
- Sehen
- Brille

Eisen: feilen = Holz:

- hobeln
- sägen
- hacken
- färben
- löten

Zeit: Maßband = Preis:

- Waage
- Meterstab
- Lineal
- Kompass
- Fernglas

Katze: Löwe = Hund:

- Wolf
- Tiger
- Bär
- Rudel
- Jaguar

Supplement 4 — Experiment 2: Main Analyses with Attitude Towards AI as Control Variable

I reconducted the main analyses reported in the manuscript with an additional consideration of the attitude towards AI as a covariate. Importantly, results did not differ when this variable was additionally considered. In all of the reported analyses, the attitude towards AI was entered as a covariate. I could replicate the results of Experiment 1 when presenting a co-working scenario with an artificial intelligence instead of a robot with embodiment. Being inferior to an artificial intelligence induced status threat, $F(1, 586) = 25.71$, $p < .001$, $\eta_p^2 = .04$. Using model 4 of the PROCESS-macro for SPSS with the attitude towards AI entered as a covariate, a significant indirect effect occurred, $B = -0.20$, bootstrapped $SE = 0.05$, bootstrapped 95% CI [−0.29, −0.12]. As such, H4 and H5 gained statistical significance, while H4 turned out significant in the other direction than hypothesized.

Afterward, I extended the approach and added perceived usefulness and the mindset about human minds in the model, using the PROCESS model 90, again considering the attitude towards AI as a covariate. Hypothesis 6 was supported by a significant indirect effect $B = -0.95$, bootstrapped $SE = 0.09$, bootstrapped 95% CI [−1.13, −0.78], so perceived usefulness also served as a mediator of the condition of willingness to interact when I controlled for the attitude towards AI. The serial mediation model turned out significant for all levels of mindsets about human minds, as all confidence intervals (Johnson-Neyman-Technique) did not include zero, leading to a confirmation of H7. This result was also confirmed by PROCESS model 6, which showed a significant indirect effect for the serial mediation without the inclusion of the moderator variable, $B = -0.09$, bootstrapped $SE = 0.02$, bootstrapped 95% CI [−0.14, −0.05]. Lastly, and again considering the mindset about human minds in the analyses by using the PROCESS model 90, I found no influence of this moderator variable on the model and no significant index of moderated mediation $B = 0.01$, bootstrapped $SE = 0.01$, bootstrapped 95% CI [−0.01, 0.03], implying that a growth mindset did not strengthen willingness to interact, and leading to a rejection of H8. I summarize that the participants' attitude towards AI, which did not differ between groups, $t(587) = 0.15$, $p = .878$, did not influence the main results of the data analysis reported in the manuscript.

6. General Discussion

Artificial intelligence transforms economies and societies to such an extent that there is a debate about a fourth industrial revolution. The fourth stage builds on the progress made in the previous stages (Philbeck & Davis, 2018; Xu et al., 2018): While the first industrial revolution became possible with the proliferation of steam machines aiming to mechanize production, the second industrial revolution concentrated on electricity and telecommunication systems as well as mass production. Digital information technologies and automation symbolize the third industrial revolution. Building on these progresses, the current fourth industrial revolution is characterized by the rapid exchange of information as well as technologies to be seamlessly embedded into our physical environment. This advancement is made possible by the fusion of technologies and the widespread use of artificial intelligence (Xu et al., 2018). The German government strongly wants to support technological progress, emphasizing the opportunities that artificial intelligence brings for industry 4.0 and the economy as a whole, and invested more than half a billion euros in a respective research agenda (Bundesministerium für Bildung und Forschung, 2016). In November 2022, for example, the funding for four AI service centers was launched to advance AI research in Germany and promote its transfer into practice (Bundesministerium für Bildung und Forschung, 2022).

Although it is widely held that industrialized countries need to invest in their international competitiveness as technology locations, the opening quote from Stephen Hawking comes again to my mind when I read about the expenditures mentioned above. It is not only necessary to invest millions of euros in the development of artificial intelligence but also to “stop for a moment and focus not only on making our artificial intelligence better and more successful but also on the benefit of humanity” (Koestier, 2017). It is intelligible to invest in the technical progress of artificial intelligence and encourage open science approaches to make systems more understandable for users. However, it is also indispensable to rethink what the developments mean and include for humans who have to work, interact, and live with the new technology—opening enormous potential for discussion and research queries for psychologists, lawyers, and the humanities to make sure that progress is not only for the sake of progress but also to offer a real benefit for as many people as possible (e.g., Lugrin et al., 2022).

Based on the question of which technological improvements in the field of artificial intelligence and robots supposedly equipped with this technology may become possible and the question of whether these developments would indeed be appreciated by my study participants, I explored how people would respond to robots equipped with artificial intelligence so that they seem to be able to have their own mind. Being aware that there are several other approaches to defining and emulating the human mind, this dissertation built on the mind perception dichotomy. As such, the presented machines were described to be equipped with forms of agency and experience, which are the two central components of human mind (Gray et al., 2007) that have gained scholarly attention in human-machine interaction over the last years (e.g., Appel et al., 2020; Gray & Wegner, 2012; Taylor et al., 2020). Even if especially experience was emphasized as essential for the upcoming aversion to sophisticated machines, apart from Project 1, I deliberately decided to combine the experience and agency characteristics in most of my experiments as both factors form human minds (of healthy adults) and therefore, human-*like* minds. This way, I strived to offer a credible presentation of robots with mind instead of presenting a machine that is solely equipped with experience or solely agency.

To explore how people respond to machines equipped with human-like mental capabilities, five online experiments and one laboratory experiment were conducted in the scope of this thesis, culminating in three research projects. These were carried out to explore the uncanny valley of mind by (a) using a multi-method approach (Project 1: text vignettes, Project 2: videos, Project 3: interactions), (b) presenting the robot with or without human-like mental capabilities in concrete scenarios to create a rather realistic experience and (c) interpreting the implications of interactions with robots with mental capabilities with the help of the recorded variables.

In the following section (Chapter 6.1), I will integrate my main empirical findings: After summarizing these and the methodological observations (Chapter 6.1.1), I elaborate on the connections between my projects and derive their implications for developers (Chapter 6.1.2). Thereupon, I shortly resume my findings in light of the research goals (Chapter 6.1.3). Afterward, the limitations of the work and implications for scholars' future work are discussed (Chapter 6.2). Finally, the thesis ends with concluding remarks about future human-machine interaction (Chapter 6.3).

6.1 Integration of Main Empirical Findings

6.1.1 Summary of the Findings

Project 1, using text vignettes as stimuli, showed that a thought detector robot evokes higher eeriness than an emotion detector robot. This evidence turned out stable across the six HEXACO-personality dimensions except for an unexpected interaction effect with conscientiousness. These results reveal that people see it as crucial to be the holder of their thoughts and do not want their thoughts to be read by robots. In my reading, the results of this first project show that eeriness evoked by robots with human-like mental capabilities depends not only on a robot having the same capabilities as humans but on a robot having *even more* mental capabilities than humans. Hence, my psychological research demonstrates discrepancies between what may be once technically possible and what people appreciate to be developed further.

In Project 2, using videos as stimuli, empathy mediated the effect of the situation on likeability. Additionally, another new possibility to evoke the uncanny valley of mind was revealed (beyond thought detection in Project 1): By showing machines to be human-like by their ability to fail, aversion was evoked and not (like pre-registered) with the description of a robot's mental capabilities. Since participants rated the machine with human-like mental capabilities positively regardless of the manipulated situation, no uncanny valley of mind in neutral situations was observed. Moreover, a negative direct effect of the situation on likeability came up. This effect could be attributed to unforeseen robot vulnerability in the harmful situation. The results of this project accentuate that more variables can induce human aversion to robots than I have assumed to be important based on my prior literature review. Not only a too-human-like appearance or a too-human-like mind is evaluated negatively, but also human likeness in the ability to fail.

Project 3, offering live interactions with the robot NAO and a simulated artificial intelligence, highlighted that status threat was higher if the participant contributed less to solving the verbal-creative task than the sophisticated machine. This result entails that status threat is a variable transferable from human-human interaction to human-machine interaction and implies that status threat occurs in corresponding ways like a threat to human uniqueness does in response to machines with human-like mental capabilities. Contrary to my expectations from the uncanny valley of mind and interpersonal literature, higher status threat was associated with a higher willingness to interact with the robot and the artificial intelligence in both experiments. The perceived usefulness of the machine could explain this finding: According to the data, perceived usefulness was a positive predictor of status threat and an even stronger positive predictor of willingness to interact with the outperforming machine than status threat. As such, the machine's high usefulness was responsible for the willingness to interact with it and—at the same time—the status threat evoked by it.

Beyond these substantive findings, I offer new insights into the occurrence of the uncanny valley of mind by using several methods. In line with earlier work (Appel et al., 2020; Kang & Sundar, 2019; Shank et al., 2021; Swiderska & Küster, 2020), Project 1 used text vignettes and revealed a new possibility to evoke aversion to sophisticated machines (i.e., thought detection). In contrast to Project 1 and earlier uncanny valley of mind research based on text vignettes, visual aspects were implicitly included in the stimuli of Projects 2 and 3 through the intended methodological extension, that is, the overcoming of pure text vignette studies and the therein based addition of the visual presentation of robots to increase ecological validity. The uncanny valley of mind did not come up when video stimuli and (written or verbal) mind descriptions were used in the second project, at least in its pre-registered understanding. Project 3 offered live interactions with a robot or a simulated artificial intelligence. Status threat emerged, delivering evidence for the uncanny valley of mind. However, status threat occurred in connection with the machine's perceived usefulness, a positive predictor of the willingness to interact with a machine (Davis, 1985). Based on these empirical findings, the central prior assumption—machines with human-like mental capabilities cause aversion—cannot be supported without reservation but is highly dependent on the conceptualization of the respective study, for example, of the chosen stimuli modality (e.g., Mara et al., 2022; Randall & Sabanovic, 2023) and the considered scenario. Furthermore, the implications of aversion appeared to vary and turned out to be more multifaceted than supposed at the beginning of this research. As such, my work solely gives first indications in which scenarios machines with mental capabilities are evaluated in a respective manner, being aware that this may not be generalizable for other methods and scenarios. In other words, the results of the experiments reported in this thesis made clear that it is essential to pay increased attention to studying the uncanny valley of mind in diverse contexts, as its occurrence turned out to be dependent on the presentation mode and the considered scenario.

6.1.2 Implications of the Findings

After summarizing the main empirical findings, I now connect and interpret the findings in a bigger picture, ending this chapter with design recommendations for developers and practitioners. In Project 1, controllability can be seen as essential when people indicate their responses towards modern-day machines. In an optimistic reading, the high aversion a thought detector robot evoked in Project 1 might still be attributed to people not yet being familiar with thought detection. However, it seems even more likely that people do not appreciate having their minds read without being asked for permission or at least being informed about it. As a consequence, they may decide to stop the interaction with the machine. Therefore, especially the aversive loss of control due to a machine being potentially capable of reading humans' inner thoughts must be taken into account when it comes to human-robot interactions (Kang, 2009). In this regard, some study participants' remarks clarify that there seems to be a thin line up to which humans accept an intelligent machine (comment from participant 38 in Project 1, Experiment 2: "Robots are fascinating but I think they need to be heavily regulated to prevent them from overpowering us as a species;" comment from participant 419 in Project 1, Experiment 2: "I think robots can be useful, but I wouldn't be ok with them actually having emotions or knowing what I am thinking"). After crossing this line, they no longer feel comfortable interacting with the technology. At this stage, the users might get the impression of not being in control over the robot, but potentially the other way around—a perspective also portrayed in several fictional movies, even if this partly contradicts the current capabilities of machines (Cave & Dihal, 2019; Osawa et al., 2022). From this project, I conclude that users always want to have the final power of decision regarding what is entrusted to the machine and what is not, underlining the pivotal demand for transparency and autonomy in human-machine interaction (e.g., Dietvorst et al., 2018; Hutmacher & Appel, 2022).

Apart from the fact that people should be granted autonomy and authority in human-machine interaction, my research shows another aspect of how the acceptability of sophisticated machines can be increased. A machine's perceived usefulness can be a decisive argument for a higher willingness to interact with it. In conversations with my participants in the first experiment of Project 3, they reported that they would have no problem working with the robot NAO on the verbal-creative tasks if NAO simplified the solution for them so that they could solve the tasks faster and with less personal effort than if the robot would not have supported them. Consequently, people are open to new technologies and their benefits if some convenience accompanies this, for example, machines carrying out the larger part of the work or making tasks easier for humans. By analogy, this can be compared to gratefully using a vacuum cleaner instead of a broom—an opportunity to handle the cleaning task faster and maybe even more accurately. Moreover, people's poorer performance compared to the sophisticated machine led to a higher status threat evoked by the machine but not to a lower willingness to interact with it and status threat emerged as a consequence of the machine's usefulness. That signifies that participants wanted to interact with a machine that they had perceived to be useful and thereby even accepted being status-threatened due to the expected benefits of the technology use (see also Quadflieg et al., 2016). The potential disadvantages for people's professional status as a response to the perhaps justified status threat that could result from the well-performing machines were less likely to be noticed by my participants compared to the benefits—at least in the setting I used for my empirical investigations. Whether a machine is equipped with or without mental capabilities to serve its purpose did not seem to matter so much to the study participants as long as the technical functionalities let the machine succeed in its tasks which is finally to the advantage of the person who decides to deploy the machine for the respective task. If these functionalities were not given, the machine's usage would have had no justification, which is an essential factor, for example, for commercial use. Thus, the shown implications go beyond the pure emergence of status threat as a negative response to robots with mind but show that modern-day machines are not solely evaluated negatively (e.g., threatening) but also positively (e.g., useful). As such, they are evaluated ambivalently (e.g., Brondi et al., 2021; Stapels & Eyssel, 2022).

Undoubtedly, the generalizability of these empirical findings about robots with sophisticated capabilities has to be tested for various tasks and, maybe even more essential, in field studies at the workplace in which mental performance is in the foreground. At this point, it should also be accentuated that a high self-reported willingness to interact does not automatically mean that people show respective behavior and will start interacting with the machine. The last step of the Technology Acceptance Model (Davis, 1985), from behavioral intention to actual behavior, was not examined in the third project. Notwithstanding, as prior research revealed that the intention to use technology is closely linked with actual technology use (Bröhl et al., 2019; Dong et al., 2017), I am confident that my data can be interpreted correspondingly, particularly in the overall context of the other projects' results.

In this regard, it can be specified that the positive capability of robots to serve as a helpful tool was in the foreground for my participants, not only in Project 3 but also in Project 2. Here, a failing robot was rated with lower likeability than a robot that could adequately perform its assigned tasks. The fact that the robot was not responsible for its failure but that its vulnerability was caused by a human harming the robot was not of interest when it came to the negative evaluation of the failing harmed machine. The robot's negative evaluation emerged due to its imperfection and uselessness rather than its described mental capabilities in Project 2. In fact, the robot described as having a mind was even evaluated with higher likeability than the robot without mind. Based on the Projects 2 and 3, I can highlight that the competence and the thereupon implicitly eliciting perceived usefulness of the machines are important predictors of positive responses to machines equipped with mind, so that they are able to serve as helpful and functional technologies managing their tasks without restrictions (e.g., Chugunova & Sele, 2022; Dietvorst et al., 2015). These corresponding findings were demonstrated across Projects 2 and 3, even though the experiments are entirely different in design. In conclusion, my data suggests that competence and perceived usefulness gained by equipping a machine with mental capabilities may—at the same time—alleviate negative feelings raised by perceiving a mind in a machine.

With this integration of findings in mind, developers should clearly highlight the benefit a user can take from the usage of a machine so that the machine's usefulness is so obtrusive that potential concerns in terms of the machine's superiority might move in the background. Whether this is also a form of ethically correct advertising is a question that definitely must be addressed but is beyond the scope of this thesis. Furthermore, developers should consider the user's need for autonomy and grant transparency about the technology's functionalities. Notably, apart from communicating the advantages a user can benefit from by using a respective machine, the focus should also be put on the technical functionalities of a device and their implications for users. By explaining intelligibly what a device can and cannot do, the user can actively decide whether to use (and benefit from) it or whether the user prefers not to use it, for example, due to concerns about data security or a potential loss of control. It would be most desirable if these two goals were not mutually exclusive and usefulness went hand in hand with user autonomy.

6.1.3 Evaluation of the Research Goals

By integrating these main empirical findings, I am confident to offer new psychological insights into human responses towards machines with an emulated human-like mind that are relevant for scholars and practitioners. With regard to the stated research goals at the beginning of my work, this thesis offers valuable new insights into the uncanny valley of mind and widens the knowledge in the field. What this research contributes to current literature is (a) an extension of the uncanny valley of mind that comes up in response to thought detection as well as in response to robot vulnerability and (b) the evidence that humans primarily perceive and want to use robots as supportive tools and attach great importance to the machine's functionality. As long as people are aware of the "higher purpose" of their use of the machine, that is, its competence and resulting benefits, they seem willing to interact with them despite aversion.

As such, I expand previous knowledge on the primary occurrence of the uncanny valley of mind by accentuating two more possibilities of how aversion to modern-day machines can be evoked. Likewise, I show that the upcoming of the uncanny valley of mind and the valence of its consequences is highly dependent on the used stimuli and the considered scenario. These conclusions are substantiated by diverse methods with differing ecological validity embedded in concrete scenarios, carefully allowing me to interpret my results regarding further implications for human-robot interaction. Finally, based on the conclusions I can draw from my empirical investigations with about 2500 participants, I now want to offer suggestions for further research to expand the uncanny valley of mind literature and compensate for the limitations in my work.

6.2 Limitations and Future Work

One of my research goals was to expand the realism of approaches with which the acceptability of sophisticated entities has been investigated so far. I am aware that also in my work, some text vignettes were used as stimulus material or as its starting point. However, by stepping from text vignettes to videos and interactions, this thesis offers a varied methodological approach to increase the realism of human-robot confrontations step by step. At the same time, no longitudinal studies or field studies have been conducted within the framework of my research. Particularly the projects on empathy and status threat seem interesting to replicate as field studies and are valuable to explore over a more extended period, to elucidate how interactions and implications for empathy and willingness to interact are influenced by time and setting. Furthermore, the above-mentioned overall interpretation of my findings must be considered under the premise that all three projects used different variables and stimuli. Several limitations must be noted in light of these boundary conditions, which can inspire future research in this seminal area.

First, the deviation from text vignettes as stimuli implies that the entity's appearance might have influenced the results (Projects 2 and 3). The two robots used in my studies were classified (Rosenthal-von der Pütten & Krämer, 2014) as relatively neutral (Atlas) or even as childlike and positive (NAO). This classification suggests that these somewhat positive appearances may have interfered with the influence of the mind manipulation meant to induce aversion. After all, it seems inevitable to choose an embodiment of a robot as soon as the goal is to refrain from pure text vignette studies and to restrict participants' individual imagination of a potential robot's appearance. Therefore, I propose that future studies on the mind of robots, which include visuals also due to reasons of ecological validity, will consider a variety of robot appearances in order to be able to generalize results from the mind manipulation across different types of robots. This way, it can be distinguished which of the user responses in studies with experimental designs can be clearly traced back to the manipulation of an entity's mental capability or which confounding effects of appearance must be taken into account since visual cues are automatically and quickly processed by people (Hernández-Méndez & Muñoz-Leiva, 2015; Koć-Januchta et al., 2017; Navon, 1977). Consequently, I assume that uncanny valley and uncanny valley of mind research should be connected (see also Stein et al., 2020; Yin et al., 2021), offering numerous possibilities for future research. For example, I think that as soon as robot appearance is considered in a study design, robot gender could potentially come into play and may interfere with the ascription of mind (female gender was associated with experience, male gender with agency; Abele & Wojciszke, 2014; Eyssel & Hegel, 2012; Otterbacher & Talias, 2017).

Second, the majority of my experiments were conducted as relatively short cross-sectional online studies; solely the third project let participants interact with the robot NAO or with an artificial intelligence in experiments that took about 15 minutes. This approach is not comparable with real-life interactions over a longer time, in which lasting personal relationships between a person and a mechanical device might develop (e.g., Kidd & Breazeal, 2008; Rakhymbayeva et al., 2021). Additionally, the first experiment of Project 3 was conducted with the Wizard-of-Oz technique. That means the user had the impression of operating with a fully functioning system while the study conductor (as a hidden wizard) operated the robot to make it appear fully operational (Salber & Coutaz, 1993). Consequently, there is the possibility that participants might have noticed that they were not *really* interacting with a system embedded with mental capabilities and, though, might have doubted the realism of the study. Notwithstanding, considerable effort was put into controlling the robot to offer the highest possible validity in line with these boundary conditions. As a next step, one can strive to overcome the Wizard-of-Oz approach by using innovative machine learning programs in language generation (e.g., GPT-3 and its successors). Scholars could let participants interact with a robot that uses such a language model and is thus enabled to answer independently and well adapted to the participants' statements. Even if this seems to be a fascinating possibility for future research, experimental comparability among participants (as assured by the Wizard-of-Oz approach) must not be lost of sight either.

Third, all data used to assess the hypotheses were collected in Western countries and solely in two Western countries, the United States of America and Germany. Based on the collected socio-demographical information, the participants have enjoyed a rather high level of education and majorly self-described themselves as Caucasian. Therefore, my samples can be regarded to be WEIRD (white, educated, industrialized, rich, and democratic; Henrich et al., 2010). The mean age of participants never exceeded 41 years. Consequently, there is a sampling bias in my data which should not be ignored so that my findings may not be generalizable for people of other cultures, religions, and backgrounds. To address this limitation, for example, it could be fascinating to reconduct the experiments in Eastern cultures as research revealed differences in the assessment of robots between Western and Eastern societies (Dang & Liu, 2022b; Rau et al., 2009; Stein & Ohler, 2018). Moreover, I assume that the technological progress of a population, prior knowledge about the interaction with machines, and individual differences, are crucial when it comes to whether humans decide to appreciate a machine based on its usefulness or reject it based on feelings of threat and eeriness.

6.3 Concluding Remarks

[…] Hier sitz ich, forme Menschen
Nach meinem Bilde,
Ein Geschlecht, das mir gleich sei,
Zu leiden, zu weinen,
Zu genießen und zu freuen sich,
Und dein nicht zu achten,
Wie ich!

(English: […] Here sit I, forming mortals after my image; A race resembling me, to suffer, to weep, to enjoy, to be glad, and thee to scorn, as I!)

Prometheus, Goethe (1774/1998)

With these words at the end of Goethe’s poem, the titan Prometheus turns away from his deities, convinced that he himself knows what a good and self-determined life looks like and that this is not dependent on his relationship with higher powers. These words were written 250 years ago but last until today. They can be transferred to what humans are doing in the current decade: trying to build and program human-like devices in terms of appearance, behavior, and mental capabilities like agency and experience. My projects about machines with such sophisticated mental capabilities, each with a different variable in front and center and using different methods, consistently show that humans appreciate a machine competently fulfilling its tasks and thus being able to benefit humans by being useful (Projects 2 and 3). This functionality emerged as decisive for a positive technology evaluation.

What also seems necessary for the acceptability of machines with sophisticated mental capabilities is, at least, that humans have control and power of action over the machines so that, for example, technologies cannot autonomously decide to analyze human thoughts (Project 1). Overall, my results show that participants can rate machines with sophisticated mental capabilities quite positively within these boundary conditions and that the emergence of aversion in response to machines with sophisticated mental capabilities, when investigated in context, is less generalizable than was evident based on the literature review at the beginning of my research. Therefore, I suggest that the uncanny valley of mind should primarily be researched in combination with a concrete scenario.

As a last remark based on my empirical insights, I conclude that humans are likely to react quite positively to future developments in the field of artificial intelligence—as long as (a) the perceived usefulness and competence of the machine is so high that people are interested in benefiting from the use of the technology, and (b) a human is the last link in the chain so that a human’s autonomy and desire for control should always take precedence over that of the machine. Therefore, future technological and scientific advances should be made to balance emphasizing the usefulness gained by embedding a machine with mental capabilities and keeping aversion evoked by such machines as low as possible by highlighting human autonomy. Along these lines, the development of artificial intelligence and robotics can have a promising future ahead of it, which also does not lose sight of the benefit for humanity either.

References

Abele, A. E., & Wojciszke, B. (2014). Communal and agentic content in social cognition: A dual perspective model. *Advances in Experimental Social Psychology*, *50*, 195-255. https://doi.org/10.1016/B978-0-12-800284-1.00004-7

Abele, A. E., Cuddy, A. J., Judd, C. M., & Yzerbyt, V. Y. (2008). Fundamental dimensions of social judgment. *European Journal of Social Psychology*, *38*(7), 1063-1065. https://doi.org/10.1002/ejsp.574

Abele, A. E., Hauke, N., Peters, K., Louvet, E., Szymkow, A., & Duan, Y. (2016). Facets of the fundamental content dimensions: Agency with competence and assertiveness—Communion with warmth and morality. *Frontiers in Psychology*, *7,* Article 1810. https://doi.org/10.3389/fpsyg.2016.01810

Adolphs, R. (2009). The social brain: Neural basis of social knowledge. *Annual Review of Psychology*, *60*, 693-716. https://doi.org/10.1146/annurev.psych.60.110707.163514

Affectiva. (2018). *Solutions.* https://www.affectiva.com/what/products/

Aguinis, H., & Bradley, K. J. (2014). Best practice recommendations for designing and implementing experimental vignette methodology studies. *Organizational Research Methods*, *17*(4), 351-371. https://doi.org/10.1177/1094428114547952

Aiken, L. S., & West, S. G. (1991). *Multiple regression: Testing and interpreting interactions.* Sage.

Alhashmi, S. F., Alshurideh, M., Kurdi, B. A., & Salloum, S. A. (2020). A systematic review of the factors affecting the artificial intelligence implementation in the health care sector. *Proceedings of the International Conference on Artificial Intelligence and Computer Vision*, 27-49. https://doi.org/10.1007/978-3-030-44289-7

Alonso-Martin, F., Malfaz, M., Sequeira, J., Gorostiza, J. F., & Salichs, M. A. (2013). A multimodal emotion detection system during human–robot interaction. *Sensors, 13*(11), 15549-15581. https://doi.org/10.3390/s131115549

Al-Subari, S. N., Zabri, S. M., & Ahmad, K. (2018). Factors influencing online banking adoption: The case of academicians in Malaysian technical university network (MTUN). *Advanced Science Letters*, *24*(5), 3193-3197. https://doi.org/10.1166/asl.2018.11342

Anderson, C., Srivastava, S., Beer, J. S., Spataro, S. E., & Chatman, J. A. (2006). Knowing your place: Self-perceptions of status in face-to-face groups. *Journal of Personality and Social Psychology*, *91,* 1094-1110. https://doi.org/10.1037/0022-3514.91.6.1094

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

Appel, M., Marker, C., & Mara, M. (2019). Otakuism and the appeal of sex robots. *Frontiers in Psychology, 10,* Article 569. https://doi.org/10.3389/fpsyg.2019.00569

Ashton, M. C., & Lee, K. (2009). The HEXACO-60: A short measure of the major dimensions of personality. *Journal of Personality Assessment, 91*, 340-345. https://doi.org/10.1080/00223890902935878

Ashton, M. C., Lee, K., & Goldberg, L. R. (2004). A hierarchical analysis of 1,710 English personality-descriptive adjectives. *Journal of Personality and Social Psychology, 87*(5), 707-721. https://doi.org/10.1037/0022-3514.87.5.707

Atzmüller, C., & Steiner, P. M. (2010). Experimental vignette studies in survey research. *Methodology, 6*(3), 128-138. https://doi.org/10.1027/1614-2241/a000014

Bakan, D. (1966). *The duality of human existence: An essay on psychology and religion*. Rand McNally.

Bakpayev, M., Baek, T. H., van Esch, P., & Yoon, S. (2022). Programmatic creative: AI can think but it cannot feel. *Australasian Marketing Journal, 30*(1), 90-95. https://doi.org/10.1016/j.ausmj.2020.04.002

Baltrusch, S. J., Krause, F., de Vries, A. W., van Dijk, W., & de Looze, M. P. (2022). What about the human in human robot collaboration? A literature review on HRC's effects on aspects of job quality. *Ergonomics, 65*(5), 719-740. https://doi.org/10.1080/00140139.2021.1984585

Bankins, S., & Formosa, P. (2020). When AI meets PC: Exploring the implications of workplace social robots and a human-robot psychological contract. *European Journal of Work and Organizational Psychology, 29*(2), 215-229. https://doi.org/10.1080/1359432X.2019.1620328

Banks, J. (2019). Theory of mind in social robots: Replication of five established human tests. *International Journal of Social Robotics, 12,* 403-414. https://doi.org/10.1007/s12369-019-00588-x

Bartneck, C., Kanda, T., Ishiguro, H., & Hagita, N. (2007, August 26-29). Is the uncanny valley an uncanny cliff? *Proceedings of the 16th IEEE International Symposium on Robot and Human Interactive Communication,* 368-373. https://doi.org/10.1109/ROMAN.2007.4415111

Bartneck, C., Kanda, T., Mubin, O., & Al Mahmud, A. (2007, November 29-December 1). The perception of animacy and intelligence based on a robot's embodiment. *Proceedings of the 7th IEEE-RAS International Conference on Humanoid Robots,* 300-305. https://doi.org/10.1109/ICHR.2007.4813884

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

Batson, C. D., & Ahmad, N. Y. (2009). Using empathy to improve intergroup attitudes and relations. *Social Issues and Policy Review, 3*(1), 141-177. https://doi.org/10.1111/j.1751-2409.2009.01013.x

Batson, C. D., Polycarpou, M. P., Harmon-Jones, E., Imhoff, H. J., Mitchener, E. C., Bednar, L. L., Klein, T. R., & Highberger, L. (1997). Empathy and attitudes: Can feeling for a member of a stigmatized group improve feelings toward the group? *Journal of Personality and Social Psychology, 72*(1), 105-118. https://doi.org/10.1037/0022-3514.72.1.105

Benishek, L. E., & Lazzara, E. H. (2019). Teams in a new era: Some considerations and implications. *Frontiers in Psychology, 10,* Article 1006. https://doi.org/10.3389/fpsyg.2019.01006

Benjamin, R., & Heine, S. J. (2023). From Freud to android: Constructing a scale of uncanny feelings. *Journal of Personality Assessment, 105*(1), 121-133. https://doi.org/10.1080/00223891.2022.2048842

Bianco, F., & Ognibene, D. (2019). Transferring adaptive theory of mind to social robots: Insights from developmental psychology to robotics. In M. A. Salichs, S. S. Ge, E. I. Barakova, J.-J. Cabibihan, A. R. Wagner, A. Castro-González, & H. He (Eds.), *ICSR 2019: Social robotics* (pp. 77-78). Springer. https://doi.org/10.1007/978-3-030-35888-4

Bigman, Y. E., & Gray, K. (2018). People are averse to machines making moral decisions. *Cognition, 181*, 21-34. https://doi.org/10.1016/j.cognition.2018.08.003

Bigman, Y. E., Yam, K. C., Marciano, D., Reynolds, S. J., & Gray, K. (2021). Threat of racial and economic inequality increases preference for algorithm decision-making. *Computers in Human Behavior, 122*, Article 106859. https://doi.org/10.1016/j.chb.2021.106859

Binder, J., Zagefka, H., Brown, R., Funke, F., Kessler, T., Mummendey, A., Maquil, A., Demoulin, S., & Leyens, J.-P. (2009). Does contact reduce prejudice or does prejudice reduce contact? A longitudinal test of the contact hypothesis among majority and minority groups in three European countries. *Journal of Personality and Social Psychology, 96*(4), 843-856. https://doi.org/10.1037/a0013470

Bodala, I. P., Churamani, N., & Gunes, H. (2020). Creating a robot coach for mindfulness and wellbeing: A longitudinal study. *arXiv preprint.* Advance online publication. https://doi.org/10.48550/arXiv.2006.05289

Borenstein, J. (2011). Robots and the changing workforce. *AI & Society, 26*(1), 87-93. https://doi.org/10.1007/s00146-009-0227-0

Boroumand, S., Eys, M., & Benson, A. J. (2018). How status conflict undermines athletes' willingness to help new teammates. *Journal of Applied Sport Psychology, 30*(3), 358-365. https://doi.org/10.1080/10413200.2017.1384939

Bothner, M., Kang, J., & Stuart, T. (2007). Competitive crowding and risk taking in a tournament: Evidence from NASCAR racing. *Administrative Science Quarterly, 52*, 208-247. http://dx.doi.org/10.2189/asqu.52.2.208

Breazeal, C., Gray, J., & Berlin, M. (2009). An embodied cognition approach to mindreading skills for socially intelligent robots. *The International Journal of Robotics Research, 28*(5), 656-680. https://doi.org/10.1177/0278364909102796

Breazeal, C., Harris, P. L., DeSteno, D., Kory Westlund, J. M., Dickens, L., & Jeong, S. (2016). Young children treat robots as informants. *Topics in Cognitive Science, 8*(2), 481-491. https://doi.org/10.1111/tops.12192

Brewer, M. B. (2000). Research design and issues of validity. In H. T. Reis & C. M. Judd (Eds.), *Handbook of research methods in social and personality psychology* (pp. 3–16). Cambridge University Press.

Breza, E., Kaur, S., & Shamdasani, Y. (2018). The morale effects of pay inequality. *The Quarterly Journal of Economics, 133*(2), 611-663. https://doi.org/10.1093/qje/qjx041

Brink, K. A., Gray, K., & Wellman, H. M. (2019). Creepiness creeps in: Uncanny valley feelings are acquired in childhood. *Child Development, 90*(4), 1202-1214. https://doi.org/10.1111/cdev.12999

Brislin, R. W. (1970). Back-translation for cross-cultural research. *Journal of Cross-Cultural Psychology, 1*(3), 185-216. https://doi.org/10.1177/135910457000100301

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

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*, *11*(5), 709-726. https://doi.org/10.1007/s12369-019-00593-0

Broman, M. M., & Finckenberg-Broman, P. (2017, August 10-11). Human-robotics & AI interaction: The Robotics/AI legal entity (RAiLE©). *Proceedings of the IEEE International Symposium on Technology and Society (ISTAS),* 1-7. https://doi.org/10.1109/ISTAS.2017.8318980

Brondi, S., Pivetti, M., Di Battista, S., & Sarrica, M. (2021). What do we expect from robots? Social representations, attitudes and evaluations of robots in daily life. *Technology in Society, 66,* Article 101663. https://doi.org/10.1016/j.techsoc.2021.101663

Brooks, A. W., Dai, H., & Schweitzer, M. E. (2014). I'm sorry about the rain! Superfluous apologies demonstrate empathic concern and increase trust. *Social Psychological and Personality Science*, *5*(4), 467-474. https://doi.org/10.1177/1948550613506122

Brooks, C., & Szafir, D. (2019, November 7-9). Building second-order mental models for human-robot interaction. In *2019 AAAI Fall Symposium* [Symposium]. Association for the Advancement of Artificial Intelligence, Washington, DC, United States. https://doi.org/10.48550/arXiv.1909.06508

Brooks, D. J., Begum, M., & Yanco, H. A. (2016, August 26-31). Analysis of reactions towards failures and recovery strategies for autonomous robots. *Proceedings of the 25th IEEE International Symposium on Robot and Human Interactive Communication*, 487-492. https://doi.org/10.1109/ROMAN.2016.7745162

Brščić, D., Kidokoro, H., Suehiro, Y., & Kanda, T. (2015, March 2-5). Escaping from children's abuse of social robots. *Proceedings of the 10th Annual ACM/IEEE International Conference on Human-Robot Interaction,* 59-66. https://doi.org/10.1145/2696454.2696468

Bruneau, E. G., & Saxe, R. (2012). The power of being heard: The benefits of 'perspective-giving' in the context of intergroup conflict. *Journal of Experimental Social Psychology*, *48*(4), 855-866. https://doi.org/10.1016/j.jesp.2012.02.017

Bryndin, E. (2020). Formation of technological cognitive reason with artificial intelligence in virtual space. *Britain International of Exact Sciences (BIoEx) Journal*, *2*(2), 450-461. https://doi.org/10.33258/bioex.v2i2.222

Bughin, J., Seong, J., Manyika, J., Chui, M., & Joshi, R. (2018, September 4). Notes from the AI frontier: Modeling the impact of AI on the world economy. *McKinsey.* https://mck.co/3OLwfkD

Bundesministerium für Bildung und Forschung. (2016). *Industrie 4.0.* https://www.bmbf.de/bmbf/de/forschung/digitale-wirtschaft-und-gesellschaft/industrie-4-0/industrie-4-0.html

Bundesministerium für Bildung und Forschung. (2022). *Förderung von vier KI-Servicezentren gestartet.* https://www.bmbf.de/bmbf/shareddocs/kurzmeldungen/de/2022/11/foerderung-von-4-ki-zentren-gestartet.html

Buser, T., & Dreber, A. (2016). The flipside of comparative payment schemes. *Management Science*, *62*(9), 2626-2638. https://doi.org/10.1287/mnsc.2015.2257

Call, J., & Tomasello, M. (2008). Does the chimpanzee have a theory of mind? 30 years later. *Trends in Cognitive Sciences*, *12*(5), 187-192. https://doi.org/10.30965/9783957438843_008

Cameron, D., de Saille, S., Collins, E. C., Aitken, J. M., Cheung, H., Chua, A., Loh, E. L., & Law, J. (2021). The effect of social-cognitive recovery strategies on likability, capability and trust in social robots. *Computers in Human Behavior*, *114*, Article 106561. https://doi.org/10.1016/j.chb.2020.106561

Campbell, E. M., Liao, H., Chuang, A., Zhou, J., & Dong, Y. (2017). Hot shots and cool reception? An expanded view of social consequences for high performers. *Journal of Applied Psychology*, *102*(5), 845-866. https://doi.org/10.1037/apl0000183

Campbell, M., Hoane, A. J., & Hsu, F. H. (2002). Deep blue. *Artificial Intelligence*, *134*(1-2), 57-83. https://doi.org/10.1016/S0004-3702(01)00129-1

Cao, X. (2013). The effects of facial close-ups and viewers’ sex on empathy and intentions to help people in need. *Mass Communication and Society*, *16*(2), 161-178. https://doi.org/10.1080/15205436.2012.683928

Cave, S., & Dihal, K. (2019). Hopes and fears for intelligent machines in fiction and reality. *Nature Machine Intelligence*, *1*, 74-78. https://doi.org/10.1038/s42256-019-0020-9

Chang, W., Wang, H., Yan, G., Lu, Z., Liu, C., & Hua, C. (2021). EEG based functional connectivity analysis of human pain empathy towards humans and robots. *Neuropsychologia*, *151*, Article 107695. https://doi.org/10.1016/j.neuropsychologia.2020.107695

Chen, L., Su, W., Feng, Y., Wu, M., She, J., & Hirota, K. (2020). Two-layer fuzzy multiple random forest for speech emotion recognition in human-robot interaction. *Information Sciences*, *509*, 150-163. https://doi.org/10.1016/j.ins.2019.09.005

Chen, N., Mohanty, S., Jiao, J., & Fan, X. (2021). To err is human: Tolerate humans instead of machines in service failure. *Journal of Retailing and Consumer Services*, *59*, Article 102363. https://doi.org/10.1016/j.jretconser.2020.102363

Cheng, J. T., Tracy, J. L., & Henrich, J. (2010). Pride, personality, and the evolutionary foundations of human social status. *Evolution and Human Behavior*, *31*(5), 334-347. https://doi.org/10.1016/j.evolhumbehav.2010.02.004

Chien, S. Y., Sycara, K., Liu, J. S., & Kumru, A. (2016). Relation between trust attitudes toward automation, Hofstede’s cultural dimensions, and big five personality traits. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *60*(1), 841-845. https://doi.org/10.1177/1541931213601192

Choi, S., Mattila, A. S., & Bolton, L. E. (2021). To err is human (-oid): How do consumers react to robot service failure and recovery? *Journal of Service Research*, *24*(3), 357-371. https://doi.org/10.1177/1094670520978798

Chounta, I. A., Bardone, E., Raudsep, A., & Pedaste, M. (2021). Exploring teachers' perceptions of artificial intelligence as a tool to support their practice in Estonian K-12 education. *International Journal of Artificial Intelligence in Education*, *32,* 725-755. https://doi.org/10.1007/s40593-021-00243-5

Chugunova, M., & Sele, D. (2022). We and it: An interdisciplinary review of the experimental evidence on human-machine interaction. *Journal of Behavioral and Experimental Economics, 99,* Article 101897. https://doi.org/10.1016/j.socec.2022.101897

Cicero. (1977). *Cicero: Selected political speeches* (M. Grant, Trans.). Penguin Classics. (Original work published ca. 52 B.C.E.)

Cikara, M., Bruneau, E. G., & Saxe, R. R. (2011). Us and them: Intergroup failures of empathy. *Current Directions in Psychological Science, 20*(3), 149-153. https://doi.org/10.1177/0963721411408713

Cohen-Charash, Y. (2009). Episodic envy. *Journal of Applied Social Psychology, 39,* 2128–2173. http://dx.doi.org/10.1111/j.1559-1816.2009.00519.x

Colleoni, E., Rozza, A., & Arvidsson, A. (2014). Echo chamber or public sphere? Predicting political orientation and measuring political homophily in twitter using big data. *Journal of Communication, 64*(2), 317-332. https://doi.org/10.1111/jcom.12084

Conti, D., Commodari, E., & Buono, S. (2017). Personality factors and acceptability of socially assistive robotics in teachers with and without specialized training for children with disability. *Life Span and Disability, 20*(2), 251-272. http://shura.shu.ac.uk/id/eprint/18254

Craig, K., Thatcher, J. B., & Grover, V. (2019). The IT identity threat: A conceptual definition and operational measure. *Journal of Management Information Systems, 36*(1), 259-288. https://doi.org/10.1080/07421222.2018.1550561

Cuddy, A. J., Fiske, S. T., Kwan, V. S., Glick, P., Demoulin, S., Leyens, J. P., Bond, M. H., Croizet, J.-C., Ellemers, N., Sleebos, E., Htun, T. T., Kim, H.-J., Maio, G., Perry, J., Petkova, K., Todorov, V., Rodriguez-Ballon, R., Morales, E., Moya, M., Palacios, M., Smith, V., Perez, R., Vala, J., & Ziegler, R. (2009). Stereotype content model across cultures: Towards universal similarities and some differences. *British Journal of Social Psychology, 48*(1), 1-33. https://doi.org/10.1348/014466608X314935

Cuff, B. M., Brown, S. J., Taylor, L., & Howat, D. J. (2016). Empathy: A review of the concept. *Emotion Review, 8*(2), 144-153. https://doi.org/10.1177/1754073914558466

Cummins, L. F., Nadorff, M. R., & Kelly, A. E. (2009). Winning and positive affect can lead to reckless gambling. *Psychology of Addictive Behaviors, 23*(2), 287-294. https://doi.org/10.1037/a0014783

D'Cruz, P., & Noronha, E. (2021). Workplace bullying in the context of robotization: Contemplating the future of the field. In P. D'Cruz, E. Noronha, G. Notelaers, & C. Rayner (Eds.), *Handbook of workplace bullying, emotional abuse and harassment: Concepts, approaches and methods* (pp. 293-322). Springer.

Da Xu, L., Lu, Y., & Li, L. (2021). Embedding blockchain technology into IoT for security: A survey. *IEEE Internet of Things Journal, 8*(13), 10452-10473. https://doi.org/10.1109/JIOT.2021.3060508

Dai, H. (2018). A double-edged sword: How and why resetting performance metrics affects motivation and performance. *Organizational Behavior and Human Decision Processes, 148*, 12-29. https://doi.org/10.1016/j.obhdp.2018.06.002

Dang, J., & Liu, L. (2021). Robots are friends as well as foes: Ambivalent attitudes toward mindful and mindless AI robots in the United States and China. *Computers in Human Behavior*, *115*, 1-8. https://doi.org/10.1016/j.chb.2020.106612

Dang, J., & Liu, L. (2022a). A growth mindset about human minds promotes positive responses to intelligent technology. *Cognition*, *220*, Article 104985. https://doi.org/10.1016/j.cognition.2021.104985

Dang, J., & Liu, L. (2022b). Implicit theories of the human mind predict competitive and cooperative responses to AI robots. *Computers in Human Behavior*, *134*, Article 107300. https://doi.org/10.1016/j.chb.2022.107300

Darling, K. (2015, August 19). *Robot ethics is about humans* [Video]. The conference. http://videos.theconference.se/robots-and-humans

Darling, K., Nandy, P., & Breazeal, C. (2015, August 31-September 4). Empathic concern and the effect of stories in human-robot interaction. *Proceedings of the 24th IEEE International Symposium on Robot and Human Interactive Communication*, 770-775. https://doi.org/10.1109/ROMAN.2015.7333675

Darwin, C. (2009). *The expression of the emotions in man and animals* (P. Ekman, Ed.). Oxford University Press. (Original work published 1872)

Davis, F. D. (1985). *A technology acceptance model for empirically testing new end-user information systems: Theory and results* [Doctoral dissertation, Massachusetts Institute of Technology]. Massachusetts Institute of Technology.

Davis, F. D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Quarterly*, *13*(3), 319-340. https://doi.org/10.2307/249008

Davis, F. D., Bagozzi, R. P., & Warshaw, P. R. (1989). User acceptance of computer technology: A comparison of two theoretical models. *Management Science, 35*(8), 982-1003. https://doi.org/10.1287/mnsc.35.8.982

de Jong, D., Hortensius, R., Hsieh, T. Y., & Cross, E. S. (2021). Empathy and schadenfreude in human–robot teams. *Journal of Cognition*, *4*(1), Article 35. https://doi.org/10.5334/joc.177

de Melo, C., Marsella, S., & Gratch, J. (2016). People do not feel guilty about exploiting machines. *ACM Transactions on Computer-Human Interaction, 23*(2), 1-17. https://doi.org/10.1145/2890495

de Vignemont, F., & Singer, T. (2006). The empathic brain: How, when and why? *Trends in Cognitive Sciences, 10*(10), 435-441. https://doi.org/10.1016/j.tics.2006.08.008

Diel, A., & MacDorman, K. F. (2021). Creepy cats and strange high houses: Support for configural processing in testing predictions of nine uncanny valley theories. *Journal of Vision*, *21*(4), 1-20. https://doi.org/10.1167/jov.21.4.1

Diel, A., Weigelt, S., & MacDorman, K. F. (2022). A meta-analysis of the uncanny valley's independent and dependent variables. *ACM Transactions on Human-Robot Interaction (THRI)*, *11*(1), 1-33. https://doi.org/10.1145/3470742

Dietvorst, B. J., Simmons, J. P., & Massey, C. (2015). Algorithm aversion: People erroneously avoid algorithms after seeing them err. *Journal of Experimental Psychology: General, 144,* 114-126. http://dx.doi.org/10.1037/xge0000033

Dietvorst, B. J., Simmons, J. P., & Massey, C. (2018). Overcoming algorithm aversion: People will use imperfect algorithms if they can (even slightly) modify them. *Management Science*, *64*(3), 1155-1170. https://doi.org/10.1287/mnsc.2016.2643

Dissing, L., & Bolander, T. (2020, July 11-17). Implementing theory of mind on a robot using dynamic epistemic logic. *Proceedings of the 29th International Joint Conference on Artificial Intelligence,* 1615-1621. https://doi.org/10.24963/ijcai.2020/224

Dong, X., Chang, Y., Wang, Y., & Yan, J. (2017). Understanding usage of Internet of Things (IOT) systems in China: Cognitive experience and affect experience as moderator. *Information Technology & People, 30*(1), 117-138. https://doi.org/10.1108/ITP-11-2015-0272

Duan, L., Xu, L., Liu, Y., & Lee, J. (2009). Cluster-based outlier detection. *Annals of Operations Research, 168*(1), 151-168. https://doi.org/10.1007/s10479-008-0371-9

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

Duffy, M. K., Scott, K. L., Shaw, J. D., Tepper, B. J., & Aquino, K. (2012). A social context model of envy and social undermining. *Academy of Management Journal, 55,* 643-666. http://dx.doi.org/10.5465/amj.2009.0804

Dunn, J. R., & Schweitzer, M. E. (2006). Green and mean: Envy and social undermining in organizations. In E. Salas (Ed.), *Research on managing groups and teams* (pp. 177-197). Emerald.

Edgington, E. S. (1996). Randomized single-subject experimental designs. *Behaviour Research and Therapy, 34,* 567-574. https://doi.org/10.1016/00057967(96)00012-5

Ekerim-Akbulut, M., Selçuk, B., Slaughter, V., Hunter, J. A., & Ruffman, T. (2020). In two minds: Similarity, threat, and prejudice contribute to worse mindreading of outgroups compared with an ingroup. *Journal of Cross-Cultural Psychology, 51*(1), 25-48. https://doi.org/10.1177/0022022119883699

Eklund, J. H., & Meranius, M. S. (2021). Toward a consensus on the nature of empathy: A review of reviews. *Patient Education and Counseling, 104*(2), 300-307. https://doi.org/10.1016/j.pec.2020.08.022

Elliott, M. N., McCaffrey, D. F., & Lockwood, J. R. (2007). How important is exact balance in treatment and control sample sizes to evaluations? *Journal of Substance Abuse Treatment, 33*(1), 107-110. https://doi.org/10.1016/j.jsat.2006.12.007

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

Erikson, E. H. (1950). *Childhood and society*. W. W. Norton & Co.

Esterwood, C., & Robert, L. P. (2020, November 10-13). Personality in healthcare human robot interaction (H-HRI): A literature review and brief critique. *Proceedings of the 8th International Conference on Human-Agent Interaction*, 87-95. https://doi.org/10.1145/3406499.3415075

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., & Bobinger, S. (2011, March 6-9). Effects of anticipated human-robot interaction and predictability of robot behavior on perceptions of anthropomorphism. *Proceedings of the 6th International Conference on Human-Robot Interaction*, 61-68. https://doi.org/10.1145/1957656.1957673

Eyssel, F., Kuchenbrandt, D., Bobinger, S., De Ruiter, L., & Hegel, F. (2012, March 5-8). 'If you sound like me, you must be more human': On the interplay of robot and user features on human-robot acceptance and anthropomorphism. *Proceedings of the 7th annual ACM/IEEE International Conference on Human-Robot Interaction*, 125-126. https://doi.org/10.1145/2157689.2157717

Fallatah, A., Urann, J., & Knight, H. (2019, November 3-8). The robot show must go on: Effective responses to robot failures. *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 325-332. https://doi.org/10.1109/IROS40897.2019.8967854

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, *39*(2), 175-191. https://doi.org/10.3758/BF03193146

Fayard, J. V., Roberts, B. W., Robins, R. W., & Watson, D. (2012). Uncovering the affective core of conscientiousness: The role of self-conscious emotions. *Journal of Personality, 80*(1), 1-32. https://doi.org/10.1111/j.1467-6494.2011.00720.x

Fei-Fei, L., Deng, J., Russakovsky, O., Berg, A., & Li, K. (2021). *ImageNet* (Version 2021) [Data set]. https://www.image-net.org/index.php

Ferrari, F., Paladino, M. P., & Jetten, J. (2016). Blurring human–machine distinctions: Anthropomorphic appearance in social robots as a threat to human distinctiveness. *International Journal of Social Robotics*, *8*(2), 287-302. https://doi.org/10.1007/s12369-016-0338-y

Ferron, J. M., Moeyaert, M., van den Noortgate, W., & Beretvas, S. N. (2014). Estimating causal effects from multiple-baseline studies: Implications for design and analysis. *Psychological Methods, 19*(4), 493-510. https://doi.org/10.1037/a0037038

Festinger, L. (1954). A theory of social comparison processes. *Human Relations, 7*(2), 117-140. http://dx.doi.org/10.1177/001872675400700202

Fiske, S. T., Cuddy, A. J. C., Glick, P., & Xu, J. (2002). A model of stereotype content as often mixed: Separate dimensions of competence and warmth respectively follow from status and competition. *Journal of Personality and Social Psychology, 82*(6), 878-902. http://dx.doi.org/10.1037/0022-3514.82.6.878

Floridi, L. (2020). AI and its new winter: From myths to realities. *Philosophy & Technology, 33*, 1-3. https://doi.org/10.1007/s13347-020-00396-6

Flynn, F. J., & Amanatullah, E. T. (2012). Psyched up or psyched out? The influence of coactor status on individual performance. *Organization Science, 23*, 402-415. http://dx.doi.org/10.1287/orsc.1100.0552

Fogg, B. J., & Nass, C. (1997, March 22-27). How users reciprocate to computers: An experiment that demonstrates behavior change. *CHI'97 Extended Abstracts on Human Factors in Computing Systems*, 331-332. https://doi.org/10.1145/1120212.1120419

Fraune, M. R., Šabanović, S., & Smith, E. R. (2017, August 28-September 1). Teammates first: Favoring ingroup robots over outgroup humans. *Proceedings of the 26th IEEE International Symposium on Robot and Human Interactive Communication*, 1432-1427. https://doi.org/10.1109/ROMAN.2017.8172492

Fraune, M. R., Sherrin, S., Šabanović, S., & Smith, E. R. (2019, March 11-14). Is human-robot interaction more competitive between groups than between individuals? *Proceedings of the 14th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, 104-113. https://doi.org/10.1109/HRI.2019.8673241

Freud, S. (2020). *Das Unheimliche* (O. Jahrhaus, Ed.). Reclam Verlag. (Original work published 1919)

Frick, W. (2015). When your boss wears metal pants. *Harvard Business Review*, *93*(6), 84-89. https://hbsp.harvard.edu/product/R1506F-PDF-ENG

Frith, C., & Frith, U. (2005). Theory of mind. *Current Biology*, *15*(17), R644-R645. http://dx.doi.org/10.1016/j.cub.2005.08.041

Fuller, M. (2014). The concept of the soul: Some scientific and religious perspectives. In M. Fuller (Ed.), *The concept of the soul: Scientific and religious perspectives* (pp. 1-4). Cambridge Scholars Publishing.

Gazzola, V., Rizzolatti, G., Wicker, B., & Keysers, C. (2007). The anthropomorphic brain: The mirror neuron system responds to human and robotic actions. *Neuroimage*, *35*(4), 1674-1684. https://doi.org/10.1016/j.neuroimage.2007.02.003

Gebauer, J. E., Wagner, J., Sedikides, C., & Neberich, W. (2013). Agency-communion and self-esteem relations are moderated by culture, religiosity, age, and sex: Evidence for the "self-centrality breeds self-enhancement" principle. *Journal of Personality*, *81*(3), 261-275. https://doi.org/10.1111/j.1467-6494.2012.00807.x

Geerts, J., de Wit, J., & de Rooij, A. (2021). Brainstorming with a social robot facilitator: Better than human facilitation due to reduced evaluation apprehension? *Frontiers in Robotics and AI*, *8*, Article 156. https://doi.org/10.3389/frobt.2021.657291

Gesellschaft für Konsumforschung. (2015). *Repräsentativ-Befragung zum Thema „Neid 3" durch die GfK*. Nürnberg: GfK Marktforschung.

Getchell, K. M., Carradini, S., Cardon, P. W., Fleischmann, C., Ma, H., Aritz, J., & Stapp, J. (2022). Artificial intelligence in business communication: The changing landscape of research and teaching. *Business and Professional Communication Quarterly*, *85*(1), 7-33. https://doi.org/10.1177/23294906221074311

Giner-Sorolla, R. (2018, January 24). Powering your interaction. *Approaching significance.* https://approachingblog.wordpress.com/2018/01/24/powering-your-interaction-2/

Gnambs, T., & Appel, M. (2019). Are robots becoming unpopular? Changes in attitudes towards autonomous robotic systems in Europe. *Computers in Human Behavior*, *93*, 53-61. https://doi.org/10.1016/j.chb.2018.11.045

Goethe, J. W. (1998). *Prometheus* (J. Whaley, Ed.). Northwestern University Press. (Original work published 1774)

Gompei, T., & Umemuro, H. (2015, August 31-September 4). A robot's slip of the tongue: Effect of speech error on the familiarity of a humanoid robot. *Proceedings of the 24th IEEE International Symposium on Robot and Human Interactive Communication*, 331-336. https://doi.org/10.1109/ROMAN.2015.7333630

Gonsior, B., Sosnowski, S., Mayer, C., Blume, J., Radig, B., Wollherr, D., & Kühnlenz, K. (2011, July 31-August 3). Improving aspects of empathy and subjective performance for HRI through mirroring facial expressions. *Proceedings of the 20th IEEE International Symposium on Robot and Human Interactive Communication,* 350-356. https://doi.org/10.1109/ROMAN.2011.6005294

Goštautaite, B., Luberte, I., Buciuniene, I., Stankeviciute, Z., Staniškiene, E., Trish, R., & Antonio, M. (2019, May 29-June 1). *Robots at work: How human-robot interaction changes work design* [Paper presentation]. EAWOP conference, Turin, Italy.

Goudey, A., & Bonnin, G. (2016). Must smart objects look human? Study of the impact of anthropomorphism on the acceptance of companion robots. *Recherche et Applications en Marketing*, *31*(2), 2-20. https://doi.org/10.1177/2051570716643961

Grant, A. M., & Shandell, M. S. (2022). Social motivation at work: The organizational psychology of effort for, against, and with others. *Annual Review of Psychology, 73,* 301-326. https://doi.org/10.1146/annurev-psych-060321-033406

Gray, H. M., Gray, K., & Wegner, D. M. (2007). Dimensions of mind perception. *Science, 315*(5812), 619. https://doi.org/10.1126/science.1134475

Gray, K., & Wegner, D. M. (2012). Feeling robots and human zombies: Mind perception and the uncanny valley. *Cognition, 125,* 125-130. https://doi.org/10.1016/j.cognition.2012.06.007

Gray, K., Knobe, J., Sheskin, M., Bloom, P., & Barrett, L. F. (2011). More than a body: Mind perception and the nature of objectification. *Journal of Personality and Social Psychology, 101*(6), 1207-1220. https://doi.org/10.1037/a0025883

Greenberg, J., Ashton-James, C. E., & Ashkanasy, N. M. (2007). Social comparison processes in organizations. *Organizational Behavior and Human Decision Processes, 102*(1), 22-41. https://doi.org/10.1016/j.obhdp.2006.09.006

Grover, S. L., & Brockner, J. (1989). Empathy and the relationship between attitudinal similarity and attraction. *Journal of Research in Personality*, *23*(4), 469-479. https://doi.org/10.1016/0092-6566(89)90015-9

Grundke, A. (2023a). If machines outperform humans: Status threat evoked by and willingness to interact with sophisticated machines in a work-related context. *Behaviour & Information Technology.* Advance online publication. https://doi.org/10.1080/0144929X.2023.2210688

Grundke, A. (2023b, September 6-8). *Does a Robotic Co-Worker Threaten Human Status at Work?* [Poster presentation]. 13th Conference of the Media Psychology Division of the German Society for Psychology (DGPs), Luxembourg.

Grundke, A., Stein, J.-P., & Appel, M. (2021, September 8-10). *Warning: This robot reads your mind! Evaluations of thought-detecting and emotion-detecting robots (and the influence of individual differences)* [Poster presentation]. 12th Conference of the Media Psychology Division of the German Society for Psychology (DGPs), Aachen, Germany.

Grundke, A., Stein, J.-P., & Appel, M. (2022a). Mind-reading machines: Distinct user responses to thought-detecting and emotion-detecting robots. *Technology, Mind, and Behavior, 3*(1), 1-12. https://doi.org/10.1037/tmb0000053

Grundke, A., Stein, J.-P., & Appel, M. (2022b, May 26-30). *Can empathy for a harmed robot buffer the uncanny valley of mind?* [Paper presentation]. 72th Annual Conference of the International Communication Association, Paris, France.

Grundke, A., Stein, J.-P., & Appel, M. (2022c, September 10-15). *The potential of human empathy to alleviate the uncanny valley of mind* [Paper presentation]. 52nd DGPs-Congress, Hildesheim, Germany.

Grundke, A., Stein, J.-P., & Appel, M. (2023). Improving evaluations of advanced robots by depicting them in harmful situations. *Computers in Human Behavior, 140,* Article 107565. https://doi.org/10.1016/j.chb.2022.107565

Gunkel, D. J. (2018). *Robot rights*. MIT Press.

Hafizoğlu, F. M., & Sen, S. (2019). Understanding the influences of past experience on trust in human-agent teamwork. *ACM Transactions on Internet Technology (TOIT), 19*(4), 1-22. https://doi.org/10.1145/3324300

Halabi, S., Dovidio, J. F., & Nadler, A. (2008). When and how do high status group members offer help: Effects of social dominance orientation and status threat. *Political Psychology, 29*(6), 841-858. https://doi.org/10.1111/j.1467-9221.2008.00669.x

Hall, A. K., Backonja, U., Painter, I., Cakmak, M., Sung, M., Lau, T., Thompson, H. J., & Demiris, G. (2017). Acceptance and perceived usefulness of robots to assist with activities of daily living and healthcare tasks. *Assistive Technology, 31*(3), 133-140. https://doi.org/10.1080/10400435.2017.1396565

Hampel, N., & Sassenberg, K. (2021). Needs-oriented communication results in positive attitudes towards robotic technologies among blue-collar workers perceiving low job demands. *Computers in Human Behavior Reports, 3*, Article 100086. https://doi.org/10.1016/j.chbr.2021.100086

Hanson, D. (2005, December 5). Expanding the aesthetic possibilities for humanoid robots. *Proceedings of the IEEE-RAS International Conference on Humanoid Robots*, 24-31.

Haslam, N., Bain, P., Douge, L., Lee, M., & Bastian, B. (2005). More human than you: Attributing humanness to self and others. *Journal of Personality and Social Psychology, 89*(6), 937-950. https://doi.org/10.1037/0022-3514.89.6.937

Hassabis, D., Kumaran, D., Summerfield, C., & Botvinick, M. (2017). Neuroscience-inspired artificial intelligence. *Neuron, 95*(2), 245-258. https://doi.org/10.1016/j.neuron.2017.06.011

Hasson, Y., Tamir, M., Brahms, K. S., Cohrs, J. C., & Halperin, E. (2018). Are liberals and conservatives equally motivated to feel empathy toward others? *Personality and Social Psychology Bulletin, 44*(10), 1449-1459. https://doi.org/10.1177/0146167218769867

Hayes, A. F. (2012). *Process: A versatile computational tool for observed variable moderation, mediation, and conditional process modeling* [White paper]. http://www.afhayes.com/public/process2012.pdf

Hayes, A. F. (2018). *Introduction to mediation, moderation, and conditional process analysis: A regression-based approach* (2nd ed.). The Guilford Press.

Hegel, F., Gieselmann, S., Peters, A., Holthaus, P., & Wrede, B. (2011). Towards a typology of meaningful signals and cues in social robotics. *Proceedings of the 2011 RO-MAN*, 72-78. https://doi.org/10.1109/ROMAN.2011.6005246

Hegel, F., Krach, S., Kircher, T., Wrede, B., & Sagerer, G. (2008, August 1-3). Understanding social robots: A user study on anthropomorphism. *Proceedings of the 17th IEEE International Symposium on Robot and Human Interactive Communication*, 574-579. https://doi.org/10.1109/ROMAN.2008.4600728

Heinke, M. S., & Louis, W. R. (2009). Cultural background and individualistic–collectivistic values in relation to similarity, perspective taking, and empathy. *Journal of Applied Social Psychology, 39*(11), 2570-2590. https://doi.org/10.1111/j.1559-1816.2009.00538.x

Helgeson, V. S., & Fritz, H. L. (1999). Unmitigated agency and unmitigated communion: Distinctions from agency and communion. *Journal of Research in Personality, 33*(2), 131-158. https://doi.org/10.1006/jrpe.1999.2241

Henrich, J., Heine, S., & Norenzayan, A. (2010). The weirdest people in the world? *Behavioral and Brain Sciences, 33*(2-3), 61-83. https://doi.org/10.1017/S0140525X0999152X

Hernández-Méndez, J., & Muñoz-Leiva, F. (2015). What type of online advertising is most effective for eTourism 2.0? An eye tracking study based on the characteristics of tourists. *Computers in Human Behavior, 50*, 618-625. https://doi.org/10.1016/j.chb.2015.03.017

Hilbig, B. E., & Zettler, I. (2015). When the cat's away, some mice will play: A basic trait account of dishonest behavior. *Journal of Research in Personality*, *57*, 72-88. https://doi.org/10.1016/j.jrp.2015.04.003

Hildt, E. (2019). Artificial intelligence: Does consciousness matter? *Frontiers in Psychology, 10*, Article 1535. https://doi.org/10.3389/fpsyg.2019.01535

Ho, C. C., & MacDorman, K. F. (2010). Revisiting the uncanny valley theory: Developing and validating an alternative to the Godspeed indices. *Computers in Human Behavior*, *26*(6), 1508-1518. https://doi.org/10.1016/j.chb.2010.05.015

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

Hoenen, M., Lübke, K. T., & Pause, B. M. (2016). Non-anthropomorphic robots as social entities on a neurophysiological level. *Computers in Human Behavior*, *57*, 182-186. https://doi.org/10.1016/j.chb.2015.12.034

Hoffmann, E. T. A. (2012). *Der Sandmann* (R. Drux, Ed.). Reclam Verlag. (Original work published 1816)

Horstmann, A. C., & Krämer, N. C. (2019). Great expectations? Relation of previous experiences with social robots in real life or in the media and expectancies based on qualitative and quantitative assessment. *Frontiers in Psychology*, *10*, Article 939. https://doi.org/10.3389/fpsyg.2019.00939

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), Article e0201581. https://doi.org/10.1371/journal.pone.0201581

Huang, B., Huan, Y., Xu, L. D., Zheng, L., & Zou, Z. (2019). Automated trading systems statistical and machine learning methods and hardware implementation: A survey. *Enterprise Information Systems, 13*(1), 132-144. https://doi.org/10.1080/17517575.2018.1493145

Huang, C. M., & Mutlu, B. (2013, June 24-28). Modeling and evaluating narrative gestures for humanlike robots. *Proceedings of Robotics: Science and Systems, 2,* 57-64. http://www.roboticsproceedings.org/rss09/p26.pdf

Huang, C., Cai, H., Xu, L., Xu, B., Gu, Y., & Jiang, L. (2019). Data-driven ontology generation and evolution towards intelligent service in manufacturing systems. *Future Generation Computer Systems, 101*, 197-207. https://doi.org/10.1016/j.future.2019.05.075

Huang, M. H., & Rust, R. T. (2017). Technology-driven service strategy. *Journal of the Academy of Marketing Science*, *45*(6), 906-924. https://doi.org/10.1007/s11747-017-0545-6

Huang, M. X., Li, J., Ngai, G., Leong, H. V., & Bulling, A. (2019, October 21-25). Moment-to-moment detection of internal thought during video viewing from eye vergence behavior. *Proceedings of the 27th ACM International Conference on Multimedia*, 2254-2262. https://doi.org/10.1145/3343031.3350573

Hutmacher, F., & Appel, M. (2023). The psychology of personalization in digital environments: From motivation to well-being – a theoretical integration. *Review of General Psychology, 27*(1), 26-40. https://doi.org/10.1177/10892680221105663

International Federation of Robotics. (2020). *IFR Press Conference.* https://ifr.org/downloads/press2018/Presentation_WR_2020.pdf

International Federation of Robotics. (2021). *Executive Summary World Robotics 2021 - Service Robots.* https://ifr.org/img/worldrobotics/Executive_Summary_WR_Service_Robots_2021.pdf

International Federation of Robotics. (2022). *World Robotics 2022 – Industrial Robots.* https://ifr.org/img/worldrobotics/Executive_Summary_WR_Industrial_Robots_2022.pdf

Javaras, K. N., Schaefer, S. M., van Reekum, C. M., Lapate, R. C., Greischar, L. L., Bachhuber, D. R., Love, G. D., Ryff, C. D., & Davidson, R. J. (2012). Conscientiousness predicts greater recovery from negative emotion. *Emotion, 12*(5), 875-881. https://doi.org/10.1037/a0028105

Jentsch, E. (1906/1997). On the psychology of the uncanny. *Angelaki: Journal of the Theoretical Humanities, 2*(1), 7-16. https://doi.org/10.1080/09697259708571910 (Reprinted from “Zur Psychologie des Unheimlichen“, 1906, *Psychiatrisch-Neurologische Wochenschrift, 8*[22-23], 195-198)

Johnson, J. A., Cheek, J. M., & Smither, R. (1983). The structure of empathy. *Journal of Personality and Social Psychology, 45*(6), 1299-1312. https://doi.org/10.1037/0022-3514.45.6.129

Jordan, M. I., & Mitchell, T. M. (2015). Machine learning: Trends, perspectives, and prospects. *Science, 349*(6245), 255-260. https://doi.org/10.1126/science.aaa8415

Judd, C. M., James-Hawkins, L., Yzerbyt, V., & Kashima, Y. (2005). Fundamental dimensions of social judgment: Understanding the relations between judgments of competence and warmth. *Journal of Personality and Social Psychology*, *89*(6), 899-913. http://dx.doi.org/10.1037/0022-3514.89.6.889

Kahn, P. H., Kanda, T., Ishiguro, H., Freier, N. G., Severson, R. L., Gill, B. T., Ruckert, J. H., & Shen, S. (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

Kamide, H., Kawabe, K., Shigemi, S., & Arai, T. (2013, October 27-29). Social comparison between the self and a humanoid. *Proceedings of the International Conference on Social Robotics*, 190-198. https://doi.org/10.1007/978-3-319-02675-6_19

Kamide, H., Mae, K., Shigemi, S., & Arai, T. (2012, May 14-18). A psychological scale for general impressions of humanoids. *Proceedings of the 7th IEEE International Conference on Robotics and Automation*, 49-56. https://doi.org/10.1109/ICRA.2012.6224790

Kammrath, L. K., & Dweck, C. (2006). Voicing conflict: Preferred conflict strategies among incremental and entity theorists. *Personality and Social Psychology Bulletin*, *32*(11), 1497-1508. https://doi.org/10.1177/0146167206291476

Kang, J., & Sundar, S. S. (2019, June 26-28). Social robots with a theory of mind (ToM): Are we threatened when they can read our emotions? *Proceedings of the International Symposium on Ambient Intelligence*, 80-88. https://doi.org/10.1007/978-3-030-24097-4_10

Kang, M. (2009). The ambivalent power of the robot. *Antennae, 1*(9), 47-58.

Kaplan, A., & Haenlein, M. (2019). Siri, Siri, in my hand: Who's the fairest in the land? On the interpretations, illustrations, and implications of artificial intelligence. *Business Horizons*, *62*(1), 15-25. https://doi.org/10.1016/j.bushor.2018.08.004

Kaplan, F. (2004). Who is afraid of the humanoid? Investigating cultural differences in the acceptance of robots. *International Journal of Humanoid Robotics*, *1*(3), 465-480. https://doi.org/10.1142/S0219843604000289

Kaseweter, K. A., Drwecki, B. B., & Prkachin, K. M. (2012). Racial differences in pain treatment and empathy in a Canadian sample. *Pain Research and Management*, *17*(6), 381-384. https://doi.org/10.1155/2012/803474

Kashive, N., Powale, L., & Kashive, K. (2021). Understanding user perception toward artificial intelligence (AI) enabled e-learning. *The International Journal of Information and Learning Technology*, *38*(1), 1-19. https://doi.org/10.1108/IJILT-05-2020-0090

Kätsyri, J., Förger, K., Mäkäräinen, M., & Takala, T. (2015). A review of empirical evidence on different uncanny valley hypotheses: Support for perceptual mismatch as one road to the valley of eeriness. *Frontiers in Psychology*, *6*, Article 390. https://doi.org/10.3389/fpsyg.2015.00390

Kelly, S. M. (2023, January 26). *ChatGPT passes exams from law and business schools.* CNN Business. https://edition.cnn.com/2023/01/26/tech/chatgpt-passes-exams/index.html

Kemper, T. D. (1991). Predicting emotions from social relations. *Social Psychology Quarterly, 54,* 330-342. http://dx.doi.org/10.2307/2786845

Kennedy, R., Clifford, S., Burleigh, T., Waggoner, P. D., Jewell, R., & Winter, N. J. (2020). The shape of and solutions to the MTurk quality crisis. *Political Science Research and Methods*, *8*(4), 614-629. https://doi.org/10.1017/psrm.2020.6

Kidd, C. D., & Breazeal, C. (2008, September 22-26). Robots at home: Understanding long-term human-robot interaction. *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems*, 3230-3235. https://doi.org/10.1109/IROS.2008.4651113

Kim, J., Merrill, K., & Collins, C. (2021). AI as a friend or assistant: The mediating role of perceived usefulness in social AI vs. functional AI. *Telematics and Informatics*, *64*, Article 101694. https://doi.org/10.1016/j.tele.2021.101694

Klein, K. J. K., & Hodges, S. D. (2001). Gender differences, motivation, and empathic accuracy: When it pays to understand. *Personality Social Psychology Bulletin, 27,* 720-730. http://dx.doi.org/10.1177/0146167201276007

Kleinlogel, E. P., Dietz, J., & Antonakis, J. (2018). Lucky, competent, or just a cheat? Interactive effects of honesty-humility and moral cues on cheating behavior. *Personality and Social Psychology Bulletin*, *44*(2), 158-172. https://doi.org/10.1177/0146167217733071

Knobe, J., & Prinz, J. (2008). Intuitions about consciousness: Experimental studies. *Phenomenology and the Cognitive Sciences*, *7*(1), 67-83. https://doi.org/10.1007/s11097-007-9066-y

Koć-Januchta, M., Höffler, T., Thoma, G. B., Prechtl, H., & Leutner, D. (2017). Visualizers versus verbalizers: Effects of cognitive style on learning with texts and pictures–An eye-tracking study. *Computers in Human Behavior*, *68*, 170-179. https://doi.org/10.1016/j.chb.2016.11.028

Koestier, J. (2017, November 6). Stephen Hawking issues stern warning on AI: Could be 'Worst Thing' for humanity. *Forbes.* https://www.forbes.com/sites/johnkoetsier/2017/11/06/stephen-hawking-issues-stern-warning-on-ai-could-be-worst-thing-for-humanity/?sh=4018b81253a7

Korukonda, A. R. (2007). Differences that do matter: A dialectic analysis of individual characteristics and personality dimensions contributing to computer anxiety. *Computers in Human Behavior*, *23*(4), 1921-1942. https://doi.org/10.1016/j.chb.2006.02.003

Krach, S., Hegel, F., Wrede, B., Sagerer, G., Binkofski, F., & Kircher, T. (2008). Can machines think? Interaction and perspective taking with robots investigated via fMRI. *PLoS ONE, 3*(7), Article e2597. https://doi.org/10.1371/journal.pone.0002597

Kratochwill, T. R., & Levin, J. R. (2010). Enhancing the scientific credibility of single-case intervention research: Randomization to the rescue. *Psychological Methods, 15,* 124-144. https://doi.org/10.1037/a0017736

Küster, D., & Swiderska, A. (2021). Seeing the mind of robots: Harm augments mind perception but benevolent intentions reduce dehumanisation of artificial entities in visual vignettes. *International Journal of Psychology, 56*(3), 454-465. https://doi.org/10.1002/ijop.12715

Küster, D., Swiderska, A., & Gunkel, D. (2020). I saw it on YouTube! How online videos shape perceptions of mind, morality, and fears about robots. *New Media & Society, 23*(11), 3312-3331. https://doi.org/10.1177/1461444820954199

Laban, G., George, J. N., Morrison, V., & Cross, E. S. (2021). Tell me more! Assessing interactions with social robots from speech. *Paladyn, Journal of Behavioral Robotics, 12*(1), 136-159. https://doi.org/10.1515/pjbr-2021-0011

Lachin, J. M. (1988). Properties of simple randomization in clinical trials. *Controlled Clinical Trials, 9*(4), 312-326. https://doi.org/10.1016/0197-2456(88)90046-3

Laird, J. E., Lebiere, C., & Rosenbloom, P. S. (2017). A standard model of the mind: Toward a common computational framework across artificial intelligence, cognitive science, neuroscience, and robotics. *AI Magazine, 38*(4), 13-26. https://doi.org/10.1609/aimag.v38i4.2744

Lam, C. K., van der Vegt, G. S., Walter, F., & Huang, X. (2011). Harming high performers: A social comparison perspective on interpersonal harming in work teams. *Journal of Applied Psychology, 96*(3), 588-601. https://doi.org/10.1037/a0021882

Lambert-Pandraud, R., & Laurent, G. (2010). Why do older consumers buy older brands? The role of attachment and declining innovativeness. *Journal of Marketing, 74*(5), 104-121. https://doi.org/10.1509/jmkg.74.5.104.

Latikka, R., Savela, N., Koivula, A., & Oksanen, A. (2021). Attitudes toward robots as equipment and coworkers and the impact of robot autonomy level. *International Journal of Social Robotics, 13*(7), 1747-1759. https://doi.org/10.1007/s12369-020-00743-9

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature, 521,* 436-444. https://doi.org/10.1038/nature14539

Lee, K. M., Jung, Y., Kim, J., & Kim, S. R. (2006). Are physically embodied social agents better than disembodied social agents?: The effects of physical embodiment, tactile interaction, and people’s loneliness in human–robot interaction. *International Journal of Human-Computer Studies, 64*(10), 962-973. https://doi.org/10.1016/j.ijhcs.2006.05.002

Lee, K., & Ashton, L. (2009). Scale descriptions. *The HEXACO personality inventory – revised.* http://www.hexaco.org/scaledescriptions

Li, T., Wang, L., Liu, J., Yuan, J., & Liu, P. (2022). Sharing the roads: Robot drivers (vs. human drivers) might provoke greater driving anger when they perform identical annoying driving behaviors. *International Journal of Human–Computer Interaction, 38*(4), 309-323. https://doi.org/10.1080/10447318.2021.1938392

Liepmann, D., Beauducel, A., Brocke, B., & Amthauer, R. (2007). *Intelligenz-Struktur-Test 2000 R (IST-2000 R).* Hogrefe.

Lischetzke, T., Izydorczyk, D., Hüller, C., & Appel, M. (2017). The topography of the uncanny valley and individuals’ need for structure: A nonlinear mixed effects analysis. *Journal of Research in Personality, 68*, 96-113. https://doi.org/10.1016/j.jrp.2017.02.001

Liu, J., Kong, X., Xia, F., Bai, X., Wang, L., Qing, Q., & Lee, I. (2018). Artificial intelligence in the 21st century. *IEEE Access, 6,* 34403-34421. https://doi.org/10.1109/ACCESS.2018.2819688

Liu, Y., & Liao, S. (2021, June 18-20). Would humans want to work side-by-side with autonomous robots? The effect of robot autonomy on perceived usefulness, ease of use and desire for contact. *Proceedings of the 2021 International Conference on Control and Intelligent Robotics*, 671-675. https://doi.org/10.1145/3473714.3473830

Long, F., Ye, Z., & Liu, G. (2023). Intergroup threat, knowledge of the outgroup, and willingness to purchase ingroup and outgroup products: The mediating role of intergroup emotions. *European Journal of Social Psychology, 53*(2), 268-287. https://doi.org/10.1002/ejsp.2902

Longoni, C., Bonezzi, A., & Morewedge, C. K. (2019). Resistance to medical artificial intelligence. *Journal of Consumer Research*, *46*(4), 629-650. https://doi.org/10.1093/jcr/ucz013

Looije, R., Neerincx, M. A., & Cnossen, F. (2010). Persuasive robotic assistant for health self-management of older adults: Design and evaluation of social behaviors. *International Journal of Human-Computer Studies*, *68*(6), 386-397. https://doi.org/10.1016/j.ijhcs.2009.08.007

Lotz-Schmitt, K., Siem, B., & Stürmer, S. (2017). Empathy as a motivator of dyadic helping across group boundaries: The dis-inhibiting effect of the recipient's perceived benevolence. *Group Processes & Intergroup Relations*, *20*(2), 233-259. https://doi.org/10.1177/1368430215612218

Lu, L., Cai, R., & Gursoy, D. (2019). Developing and validating a service robot integration willingness scale. *International Journal of Hospitality Management, 80*, 36-51. https://doi.org/10.1016/j.ijhm.2019.01.005

Lu, L., Xu, L., Xu, B., Li, G., & Cai, H. (2018). Fog computing approach for music cognition system based on machine learning algorithm. *IEEE Transactions on Computational Social Systems, 5*(4), 1142-1151. https:/doi.org/10.1109/TCSS.2018.2871694

Lucas, H., Poston, J., Yocum, N., Carlson, Z., & Feil-Seifer, D. (2016, August 26-31). Too big to be mistreated? Examining the role of robot size on perceptions of mistreatment. *Proceedings of the 25th IEEE International Symposium on Robot and Human Interactive Communication*, 1071-1076. https://doi.org/10.1109/ROMAN.2016.7745241

Lugrin, B., Dippold, J., & Bergmann, K. (2018, October 1-5). Social robots as a means of integration? An explorative acceptance study considering gender and non-verbal behaviour. *Proceedings of the 2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2026-2032. https://doi.org/10.1109/IROS.2018.8593818

Lugrin, B., Pelachaud, C., André, E., Aylett, R., Bickmore, T., Breazeal, C., Broekens, J., Dautenhahn, K., Gratch, J., Kopp, S., Nadel, J., Paiva, A., & Wykowska, A. (2022). Challenge discussion on socially interactive agents: Considerations on social interaction, computational architectures, evaluation, and ethics. In B. Lugrin, C. Pelachaud, & D. Traum (Eds.), *The handbook on socially interactive agents: 20 years of research on embodied conversational agents, intelligent virtual agents, and social robotics: Interactivity, platforms, application* (pp. 561-626). Association for Computing Machinery. https://doi.org/10.1145/3563659

Lyons, J. B., Nam, C. S., Jessup, S. A., Vo, T. Q., & Wynne, K. T. (2020, September 7-9). The role of individual differences as predictors of trust in autonomous security robots.

Proceedings of the 2020 IEEE International Conference on Human-Machine Systems (ICHMS), 1-5. https://doi.org/10.1109/ICHMS49158.2020.9209544

Macaskill, A., Maltby, J., & Day, L. (2002). Forgiveness of self and others and emotional empathy. *The Journal of Social Psychology, 142,* 663-665. http://dx.doi.org/10.1080/00224540209603925

MacDorman, K. F. (2006, July 26). Subjective ratings of robot video clips for human likeness, familiarity, and eeriness: An exploration of the uncanny valley. *Proceedings of the ICCS/CogSci-2006 Long Symposium: Toward Social Mechanisms of Android Science*, 26-29. https://doi.org/10.1.1.511.1867

MacDorman, K. F., & Entezari, S. O. (2015). Individual differences predict sensitivity to the uncanny valley. *Interaction Studies, 16*(2), 141-172. https://doi.org/10.1075/is.16.2.01mac

MacDorman, K. F., & Ishiguro, H. (2006). The uncanny advantage of using androids in cognitive and social science research. *Interaction Studies, 7*(3), 297-337. https://doi.org/10.1075/is.7.3.03mac

MacDorman, K. F., Green, R. D., Ho, C. C., & Koch, C. T. (2009). Too real for comfort? Uncanny responses to computer generated faces. *Computers in Human Behavior, 25*(3), 695-710. https://doi.org/10.1016/j.chb.2008.12.026

MacDorman, K. F., Srinivas, P., & Patel, H. (2013). The uncanny valley does not interfere with level 1 visual perspective taking. *Computers in Human Behavior, 29*(4), 1671-1685. https://doi.org/10.1016/j.chb.2013.01.051

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

Malhotra, D., & Liyanage, S. (2005). Long-term effects of peace workshops in protracted conflicts. *Journal of Conflict Resolution, 49*(6), 908-924. https://doi.org/10.1177/0022002705281153

Malinowska, J. K. (2021). What does it mean to empathise with a robot? *Minds and Machines, 31,* 361-376. https://doi.org/10.1007/s11023-021-09558-7

Malle, B. F. (2019). How many dimensions of mind perception really are there? *Proceedings of the 41st Annual Meeting of the Cognitive Science Society*, 2268-2274. https://research.clps.brown.edu/SocCogSci/Publications/Pubs/Malle_2019_How_Many_Dimensions.pdf

Manyika, J., Lund, S., Chui, M., Bughin, J., Woetzel, J., Batra, P., Ko, R., & Sanghvi, S. (2017). *What the future of work will mean for jobs, skills, and wages.* (McKinsey Global Institute Report). McKinsey & Company. https://www.mckinsey.com/featured-insights/future-of-work/jobs-lost-jobs-gained-what-the-future-of-work-will-mean-for-jobs-skills-and-wages

Mara, M., & Appel, M. (2015). Science fiction reduces the eeriness of android robots: A field experiment. *Computers in Human Behavior*, *48*, 156-162. https://doi.org/10.1016/j.chb.2015.01.007

Mara, M., Appel, M., & Gnambs, T. (2022). Human-like robots and the uncanny valley: A meta-analysis of user responses based on the Godspeed Scales. *Zeitschrift für Psychologie, 230*(1), 33-46. https://doi.org/10.1027/2151-2604/a000486

Mara, M., Stein, J.-P., Latoschik, M. E., Lugrin, B., Schreiner, C., Hostettler, R., & Appel, M. (2021). User responses to a humanoid robot observed in real life, virtual reality, 3D and 2D. *Frontiers in Psychology*, *12,* Article 633178. https://doi.org/10.3389/fpsyg.2021.633178

Marr, J. C., & Thau, S. (2014). Falling from great (and not so great) heights: How initial status position influences performance after status loss. *Academy of Management Journal*, *57,* 223-248. http://dx.doi.org/10.5465/amj.2011.0909

Mathur, M. B., & Reichling, D. B. (2009, March 9-13). An uncanny game of trust: Social trustworthiness of robots inferred from subtle anthropomorphic facial cues. *Proceedings of the 4th ACM/IEEE International Conference on Human-Robot Interaction (HRI),* 313-314. https://doi.org/10.1145/1514095.1514192

Mathur, M. B., & Reichling, D. B. (2016). Navigating a social world with robot partners: A quantitative cartography of the uncanny valley. *Cognition*, *146*, 22-32. https://doi.org/10.1016/j.cognition.2015.09.008

Mattiassi, A. D., Sarrica, M., Cavallo, F., & Fortunati, L. (2021). What do humans feel with mistreated humans, animals, robots, and objects? Exploring the role of cognitive empathy. *Motivation and Emotion, 45,* 543-555. https://doi.org/10.1007/s11031-021-09886-2

McAuliffe, W. H., Carter, E. C., Berhane, J., Snihur, A. C., & McCullough, M. E. (2020). Is empathy the default response to suffering? A meta-analytic evaluation of perspective taking's effect on empathic concern. *Personality and Social Psychology Review*, *24*(2), 141-162. https://doi.org/10.1177/1088868319887599

McEwan, I. (2019). *Machines like me*. Talese.

McQuay, L. (2018). Will robots duplicate or surpass us? The impact of job automation on tasks, productivity, and work. *Psychosociological Issues in Human Resource Management*, *6*(2), 86-91. http://dx.doi.org/10.22381/PIHRM6220189

Meade, A. W., & Craig, S. B. (2012). Identifying careless responses in survey data. *Psychological Methods*, *17*(3), 437-455. http://dx.doi.org/10.1037/a0028085

Mehrabian, A., Young, A. L., & Sato, S. (1988). Emotional empathy and associated individual differences. *Current Psychology*, *7*(3), 221-240. http://dx.doi.org/10.1007/BF02686670

Menne, I. M., & Schwab, F. (2018). Faces of emotion: Investigating emotional facial expressions towards a robot. *International Journal of Social Robotics*, *10*(2), 199-209. https://doi.org/10.1007/s12369-017-0447-2

Menon, T., Thompson, L., & Choi, H. S. (2006). Tainted knowledge vs. tempting knowledge: People avoid knowledge from internal rivals and seek knowledge from external rivals. *Management Science*, *52*(8), 1129-1144. https://doi.org/10.1287/mnsc.1060.0525

Messingschlager, T. V., & Appel, M. (2022). Creative artificial intelligence and narrative transportation. *Psychology of Aesthetics, Creativity, and the Arts.* Advance online publication. https://doi.org/10.1037/aca0000495

Meuwese, R., Cillessen, A. H., & Güroğlu, B. (2017). Friends in high places: A dyadic perspective on peer status as predictor of friendship quality and the mediating role of empathy and prosocial behavior. *Social Development*, *26*(3), 503-519. https://doi.org/10.1111/sode.12213

Michalos, G., Makris, S., Tsarouchi, P., Guasch, T., Kontovrakis, D., & Chryssolouris, G. (2015). Design considerations for safe human-robot collaborative workplaces. *Procedia CirP, 37,* 248-253. https://doi.org/10.1016/j.procir.2015.08.014

Microsoft Azure. (2018). *Cognitive services.* https://azure.microsoft.com/en-us/products/cognitive-services/#overview

Mirbabaie, M., Brünker, F., Möllmann Frick, N. R., & Stieglitz, S. (2022). The rise of artificial intelligence–understanding the AI identity threat at the workplace. *Electronic Markets, 32*(1), 73-99. https://doi.org/10.1007/s12525-021-00496-x

Mirnig, N., Stollnberger, G., Miksch, M., Stadler, S., Giuliani, M., & Tscheligi, M. (2017). To err is robot: How humans assess and act toward an erroneous social robot. *Frontiers in Robotics and AI, 4,* Article 21. https://doi.org/10.3389/frobt.2017.00021

Misselhorn, C. (2009). Empathy with inanimate objects and the uncanny valley. *Minds and Machines, 19*(3), 345-359. https://doi.org/10.1007/s11023-009-9158-2

Mittring, G. (2004). *Die Ermittlung der kleinsten hinreichend großen Stichprobe bei wissenschaftlichen Experimenten mit Randomisierung* [Doctoral dissertation, University of Cologne]. Networked Digital Library of Theses and Dissertations.

Moon, Y., & Nass, C. (1996). How “real” are computer personalities? Psychological responses to personality types in human-computer interaction. *Communication Research, 23*(6), 651-674. https://doi.org/10.1177/009365096023006002

Mori, M. (1970). The uncanny valley. *Energy, 7*(4), 33-35.

Mori, M., MacDorman, K. F., & Kageki, N. (2012). The uncanny valley [from the field]. *IEEE Robotics & Automation Magazine, 19*(2), 98-100. https://doi.org/10.1109/MRA.2012.2192811

Morsünbül, Ü. (2019). Human-robot interaction: How do personality traits affect attitudes towards robot? *Journal of Human Sciences, 16*(2), 499-504. https://doi.org/10.14687/jhs.v16i2.5636

Moshagen, M., Hilbig, B. E., & Zettler, I. (2018). The dark core of personality. *Psychological Review, 125*(5), 656-688. https://doi.org/10.1037/rev0000111

Moshagen, M., Thielmann, I., Hilbig, B. E., & Zettler, I. (2019). Meta-analytic investigations of the HEXACO Personality Inventory (-Revised). *Zeitschrift für Psychologie, 227,* 186-194. https://doi.org/10.1027/2151-2604/a000377

Mou, Y., Shi, C., Shen, T., & Xu, K. (2020). A systematic review of the personality of robot: Mapping its conceptualization, operationalization, contextualization and effects. *International Journal of Human–Computer Interaction, 36*(6), 591-605. https://doi.org/10.1080/10447318.2019.1663008

Moussawi, S., Koufaris, M., & Benbunan-Fich, R. (2021). How perceptions of intelligence and anthropomorphism affect adoption of personal intelligent agents. *Electronic Markets, 31,* 343-364. https://doi.org/10.1007/s12525-020-00411-w

Müller, B. C. N., Gao, X., Nijssen, S. R. R., & Damen, T. G. E. (2020). I, robot: How human appearance and mind attribution relate to the perceived danger of robots. *International Journal of Social Robotics, 13,* 691-701. https://doi.org/10.1007/s12369-020-00663-8

Müller, S. L., & Richert, A. (2018, June 26-29). The big-five personality dimensions and attitudes to-wards robots: A cross sectional study. *Proceedings of the 11th Pervasive*

Technologies Related to Assistive Environments Conference, 405-408. https://doi.org/10.1145/3197768.3203178

Müller-Abdelrazeq, S. L., Schönefeld, K., Haberstroh, M., & Hees, F. (2019). Interacting with collaborative robots—A study on attitudes and acceptance in industrial contexts. In O. Korn (Ed.), *Social robots: Technological, societal and ethical aspects of human-robot interaction* (pp. 101-117). Springer. https://doi.org/10.1007/978-3-030-17107-0_6

Mussweiler, T. (2003). Comparison processes in social judgment: Mechanisms and consequences. *Psychological Review*, *110*(3), 472-489. http://dx.doi.org/10.1037/0033-295X.110.3.472

Nass, C., & Lee, K. M. (2001). Does computer-synthesized speech manifest personality? Experimental tests of recognition, similarity-attraction, and consistency-attraction. *Journal of Experimental Psychology: Applied*, *7*(3), 171-81. https://doi.org/10.1037/1076-898X.7.3.171

Nass, C., & Moon, Y. (2002). Machines and mindlessness: Social responses to computers. *Journal of Social Issues*, *56*(1), 81-103. http://dx.doi.org/10.1111/0022-4537.00153

Nass, C., Moon, Y., & Green, N. (1997). Are machines gender neutral? Gender-stereotypic responses to computers with voices. *Journal of Applied Social Psychology*, *27*(10), 864-876. https://doi.org/10.1111/j.1559-1816.1997.tb00275.x

Navon, D. (1977). Forest before trees: The precedence of global features in visual perception. *Cognitive Psychology*, *9*(3), 353-383. https://doi.org/10.1016/0010-0285(77)90012-3

Nesse, R. M. (1990). Evolutionary explanations of emotions. *Human Nature*, *1*(3), 261-289. https://doi.org/10.1007/BF02733986

Ng, A. (2022, November 2). How AI can help counteract climate change. *The Batch.* https://www.deeplearning.ai/the-batch/how-ai-can-help-counteract-climate-change/

Niculescu, A., van Dijk, B., Nijholt, A., Li, H., & See, S. L. (2013). Making social robots more attractive: The effects of voice pitch, humor and empathy. *International Journal of Social Robotics*, *5*(2), 171-191. https://doi.org/10.1007/s12369-012-0171-x

Nieding, G., & Ohler, P. (2008). Mediennutzung und Medienwirkung bei Kindern und Jugendlichen. In B. Batinic & M. Appel (Eds.), *Medienpsychologie (Lehrbuch)* (pp. 379-402). Springer.

Nijssen, S. R., Müller, B. C., Baaren, R. B. V., & 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

Nilsson, N. J. (2010). *The quest for artificial intelligence: A history of ideas and achievements*. Cambridge University Press.

Nomura, T., Kanda, T., & Suzuki, T. (2006). Experimental investigation into influence of negative attitudes toward robots on human–robot interaction. *AI & Society*, *20*(2), 138-150. https://doi.org/10.1007/s00146-005-0012-7

Nov, O., & Ye, C. (2008, January 7-10). Personality and technology acceptance: Personal innovativeness in IT, openness and resistance to change. *Proceedings of the 41st Annual Hawaii International Conference on System Sciences*, 448-448. https://doi.org/10.1109/HICSS.2008.348

OpenAI (2022, November 30). ChatGPT: Optimizing language models for dialogue. *OpenAI.* https://openai.com/blog/chatgpt/

Osawa, H., Miyamoto, D., Hase, S., Saijo, R., Fukuchi, K., & Miyake, Y. (2022). Visions of artificial intelligence and robots in science fiction: A computational analysis. *International Journal of Social Robotics, 14,* 2123–2133. https://doi.org/10.1007/s12369-022-00876-z

Osborne-Crowley, K., Wilson, E., De Blasio, F., Wearne, T., Rushby, J., & McDonald, S. (2019). Empathy for people with similar experiences: Can the perception-action model explain empathy impairments after traumatic brain injury? *Journal of Clinical and Experimental Neuropsychology, 42*(1), 28-41. https://doi.org/10.1080/13803395.2019.1662375

Oswald, P. A. (1996). The effects of cognitive and affective perspective taking on empathic concern and altruistic helping. *The Journal of Social Psychology, 136*(5), 613-623. https://doi.org/10.1080/00224545.1996.9714045

Otterbacher, J., & Talias, M. (2017, March 6-9). S/he’s too warm/agentic! The influence of gender on uncanny reactions to robots. *Proceedings of the 12th ACM/IEEE International Conference on Human-Robot Interaction,* 214-223. http://dx.doi.org/10.1145/2909824.3020220

Ovid. (2004). *Metamorphoses* (D. Raeburn, Trans.). Penguin Classics. (Original work published ca. 8 C.E.)

Palomäki, J., Kunnari, A., Drosinou, M., Koverola, M., Lehtonen, N., Halonen, J., Repo, M., & Laakasuo, M. (2018). Evaluating the replicability of the uncanny valley effect. *Heliyon, 4*(11), Article e00939. https://doi.org/10.1016/j.heliyon.2018.e00939

Paluch, S., Tuzovic, S., Holz, H. F., Kies, A., & Jörling, M. (2022). “My colleague is a robot” – Exploring frontline employees’ willingness to work with collaborative service robots. *Journal of Service Management, 33*(2), 363-388. https://doi.org/10.1108/JOSM-11-2020-0406

Paluck, E. L. (2009). Reducing intergroup prejudice and conflict using the media: A field experiment in Rwanda. *Journal of Personality and Social Psychology, 96*(3), 574-587. https://doi.org/10.1037/a0011989

Paulhus, D. L., & John, O. P. (1998). Egoistic and moralistic biases in self-perception: The interplay of self-deceptive styles with basic traits and motives. *Journal of Personality, 66*(6), 1025-1060. https://doi.org/10.1111/1467-6494.00041

Pelau, C., Ene, I., & Pop, M. I. (2021). The impact of artificial intelligence on consumers’ identity and human skills. *Amfiteatru Economic, 23*(56), 33-45. https://doi.org/10.24818/EA/2021/56/33

Pereira, V., Hadjielias, E., Christofi, M., & Vrontis, D. (2021). A systematic literature review on the impact of artificial intelligence on workplace outcomes: A multi-process perspective. *Human Resource Management Review,* Article 100857. https://doi.org/10.1016/j.hrmr.2021.100857

Perugini, M., Gallucci, M., & Costantini, G. (2014). Safeguard power as a protection against imprecise power estimates. *Perspectives on Psychological Science, 9*(3), 319-332. https://doi.org/10.1177/1745691614528519

Petisca, S., Esteves, F., & Paiva, A. (2019, November 4-8). Cheating with robots: How at ease do they make us feel? *Proceedings of the 2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS),* 2102-2107. https://doi.org/10.1109/IROS40897.2019.8967790

Pettit, N. C., Sivanathan, N., Gladstone, E., & Marr, J. C. (2013). Rising stars and sinking ships: Consequences of status momentum. *Psychological Science*, *24,* 1579-1584. http://dx.doi.org/10.1177/0956797612473120

Pettit, N. C., Yong, K., & Spataro, S. E. (2010). Holding your place: Reactions to the prospect of status gains and losses. *Journal of Experimental Social Psychology, 46,* 396-401. http://dx.doi.org/10.1016/j.jesp.2009.12.007

Pfattheicher, S., Schindler, S., & Nockur, L. (2019). On the impact of honesty-humility and a cue of being watched on cheating behavior. *Journal of Economic Psychology, 71,* 159-174. https://doi.org/10.1016/j.joep.2018.06.004

Philbeck, T., & Davis, N. (2018). The fourth industrial revolution: Shaping new era. *Journal of International Affairs, 72*(1), 17-22. https://www.jstor.org/stable/26588339

Piwek, L., McKay, L. S., & Pollick, F. E. (2014). Empirical evaluation of the uncanny valley hypothesis fails to confirm the predicted effect of motion. *Cognition*, *130,* 271-277. https://doi.org/10.1016/j.cognition.2013.11.001

Plutchik, R. (1987). Evolutionary bases of empathy. In N. Eisenberg & J. Strayer (Eds.), *Empathy and its development* (pp. 38-46). Cambridge University Press.

Poliakoff, E., Beach, N., Best, R., Howard, T., & Gowen, E. (2013). Can looking at a hand make your skin crawl? Peering into the uncanny valley for hands. *Perception*, *42*(9), 998-1000. https://doi.org/10.1068/p7569

Premack, D., & Woodruff, G. (1978). Does the chimpanzee have a theory of mind? *Behavioral and Brain Sciences*, *1*(4), 515-526. https://doi.org/10.1017/S0140525X00076512

Quadflieg, S., Ul-Haq, I., & Mavridis, N. (2016). Now you feel it, now you don’t: How observing human-robot interactions and human-human interactions can make you feel eerie. *Interaction Studies*, *17*(2), 211-247. https://doi.org/10.1075/is.17.2.03qua

Ragni, M., Rudenko, A., Kuhnert, B., & Arras, K. O. (2016, August 26-31). Errare humanum est: Erroneous robots in human-robot interaction. *Proceedings of the 25th IEEE International Symposium on Robot and Human Interactive Communication*, 501-506. https://doi.org/10.1109/ROMAN.2016.7745164

Raisch, S., & Krakowski, S. (2021). Artificial intelligence and management: The automation–augmentation paradox. *Academy of Management Review, 46*(1), 192-210. https://doi.org/10.5465/amr.2018.0072

Rakhymbayeva, N., Amirova, A., & Sandygulova, A. (2021). A long-term engagement with a social robot for autism therapy. *Frontiers in Robotics and AI, 8,* Article 669972. https://doi.org/10.3389/frobt.2021.669972

Randall, N., & Sabanovic, S. (2023, March 13-16). A picture might be worth a thousand words, but it's not always enough to evaluate robots. *Proceedings of the 2023 ACM/IEEE International Conference on Human-Robot Interaction*, 437-445. https://doi.org/10.1145/3568162.3576970

Rau, P. P., Li, Y., & Li, D. (2009). Effects of communication style and culture on ability to accept recommendations from robots. *Computers in Human Behavior, 25*(2), 587-595. https://doi.org/10.1016/j.chb.2008.12.025

Reeves, B., & Nass, C. (1996). *The media equation–How people treat computers, television, and new media like real people and places.* CSLI Publications.

Reh, S., Tröster, C., & van Quaquebeke, N. (2018). Keeping (future) rivals down: Temporal social comparison predicts coworker social undermining via future status threat and envy. *Journal of Applied Psychology, 103*(4), 399-415. https://doi.org/10.1037/apl0000281

Riek, B. M., Mania, E. W., & Gaertner, S. L. (2006). Intergroup threat and outgroup attitudes: A meta-analytic review. *Personality and Social Psychology Review*, *10*(4), 336-353. https://doi.org/10.1207/s15327957pspr1004_4

Riek, L. D., Rabinowitch, T. C., Chakrabarti, B., & Robinson, P. (2009, March 9-13). How anthropomorphism affects empathy toward robots. *Proceedings of the 4th ACM/IEEE International Conference on Human-Robot Interaction,* 245-246. https://doi.org/10.1145/1514095.1514158

Robinette, P., Howard, A., & Wagner, A. R. (2017). Conceptualizing overtrust in robots: Why do people trust a robot that previously failed. In W. Lawless, R. Mittu, D. Sofge, & S. Russell (Eds.)*, Autonomy and artificial intelligence: A threat or savior?* (pp. 129-155). Springer. https://doi.org/10.1007/978-3-319-59719-5_6

Robinson, N. L., Connolly, J., Johnson, G. M., Kim, Y., Hides, L., & Kavanagh, D. J. (2018). Measures of incentives and confidence in using a social robot. *Science Robotics, 3*(21), Article eaat6963. https://doi.org/10.1126/scirobotics.aat6963

Rosenthal-von der Pütten, A. M., & Krämer, N. C. (2014). How design characteristics of robots determine evaluation and uncanny valley related responses. *Computers in Human Behavior*, *36*, 422-439. https://doi.org/10.1016/j.chb.2014.03.066

Rosenthal-von der Pütten, A. M., & Weiss, A. (2015). The uncanny valley phenomenon: Does it affect all of us? *Interaction Studies*, *16*, 206-214. https://doi.org/10.1075/is.16.2.07ros

Rosenthal-von der Pütten, A. M., Krämer, N. C., Hoffmann, L., Sobieraj, S., & Eimler, S. C. (2013). An experimental study on emotional reactions towards a robot. *International Journal of Social Robotics, 5,* 17-34. https://doi.org/org/10.1007/s12369-012-0173-8

Rosenthal-von der Pütten, A. M., Schulte, F. P., Eimler, S. C., Sobieraj, S., Hoffmann, L., Maderwald, S., Brand, M., & Krämer, N. C. (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

Rossi, S., Conti, D., Garramone, F., Santangelo, G., Staffa, M., Varrasi, S., & Di Nuovo, A. (2020). The role of personality factors and empathy in the acceptance and performance of a social robot for psychometric evaluations. *Robotics*, *9*(2), Article 39. https://doi.org/10.3390/robotics9020039

Rossi, S., Santangelo, G., Staffa, M., Varrasi, S., Conti, D., & Di Nuovo, A. (2018, August 27-31). Psychometric evaluation supported by a social robot: Personality factors and technology acceptance. *Proceedings of the 27th IEEE International Symposium on*

Robot and Human Interactive Communication, 802-807. https://doi.org/10.1109/ROMAN.2018.8525838

Roubroeks, M. A. J., Ham, J. R. C., & Midden, C. J. H. (2010). The dominant robot: Threatening robots cause psychological reactance, especially when they have incongruent goals. In T. Ploug, P. Hasle, & H. Oinas-Kukkonen (Eds.), *Persuasive Technology- Lecture Notes in Computer Science* (pp. 174-184). Springer. https://doi.org/10.1007/978-3-642-13226-1_18

Rueckert, L., & Naybar, N. (2008). Gender differences in empathy: The role of the right hemisphere. *Brain and Cognition, 67*, 162-167. http://dx.doi.org/10.1016/j.bandc.2008.01.002

Salber, D., & Coutaz, J. (1993, August 3-7). Applying the wizard of oz technique to the study of multimodal systems. *Proceedings of the International Conference on Human-Computer Interaction*, 219-230. https://doi.org/10.1007/3-540-57433-6_51

Salem, M., Eyssel, F., Rohlfing, K., Kopp, S., & Joublin, F. (2011). Effects of gesture on the perception of psychological anthropomorphism: A case study with a humanoid robot. In B. Mutlu, C. Bartneck, J. Ham, V. Evers, & T. Kanda (Eds.), *ICSR 2011: Social robotics* (pp. 31-41). Springer. https://doi.org/10.1007/978-3-642-25504-5_4

Salem, M., Eyssel, F., Rohlfing, K., Kopp, S., & Joublin, F. (2013). To err is human(-like): Effects of robot gesture on perceived anthropomorphism and likability. *International Journal of Social Robotics*, *5*(3), 313-323. https://doi.org/10.1007/s12369-013-0196-9

Salovey, P., & Rodin, J. (1984). Some antecedents and consequences of social-comparison jealousy. *Journal of Personality and Social Psychology, 47*(4), 780-792. https://doi.org/10.1037/0022-3514.47.4.780

Salvesen, B. (2021). Confirm you are a human: Perspectives on the uncanny valley. *International Journal for Digital Art History, 6,* 2-15. https://doi.org/10.11588/dah.2021.6.81164

Santamaria, T., & Nathan-Roberts, D. (2017). Personality measurement and design in human-robot interaction: A systematic and critical review. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 61*(1), 853-857. https://doi.org/10.1177/1541931213601686

Saucier, G., Thalmayer, A. G., Payne, D. L., Carlson, R., Sanogo, L., Ole-Kotikash, L., Church, A. T., Katigbak, S. M., Somer, O., Szarota, P., Szirmák, Z., & Zhou, X. (2014). A basic bivariate structure of personality attributes evident across nine languages. *Journal of Personality, 82*(1), 1-14. https://doi.org/10.1111/jopy.12028

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

Schaubroeck, J., & Lam, S. S. (2004). Comparing lots before and after: Promotion rejectees' invidious reactions to promotees. *Organizational Behavior and Human Decision Processes*, *94*(1), 33-47. https://doi.org/10.1016/j.obhdp.2004.01.001

Scheepers, D., Ellemers, N., & Sintemaartensdijk, N. (2009). Suffering from the possibility of status loss: Physiological responses to social identity threat in high status groups. *European Journal of Social Psychology, 39,* 1075-1092. http://dx.doi.org/10.1002/ejsp.609

Schmuckler, M. A. (2001). What is ecological validity? A dimensional analysis. *Infancy, 2*(4), 419-436. https://doi.org/10.1207/S15327078IN0204_02

Schrader, M. (Director). (2021). *I'm your man* [Film]. Letterbox Filmproduktion.

Schultz, P. W., Zelezny, L., & Dalrymple, N. J. (2000). A multinational perspective on the relation between Judeo-Christian religious beliefs and attitudes of environmental concern. *Environment and Behavior*, *32*(4), 576-591. https://doi.org/10.1177/00139160021972676

Schumann, K., Zaki, J., & Dweck, C. S. (2014). Addressing the empathy deficit: Beliefs about the malleability of empathy predict effortful responses when empathy is challenging. *Journal of Personality and Social Psychology, 107*(3), 475-493. https://doi.org/10.1037/a0036738

Seeger, A. M., Pfeiffer, J., & Heinzl, A. (2021). Texting with humanlike conversational agents: Designing for anthropomorphism. *Journal of the Association for Information Systems*, *22*(4), 931-967. https://doi.org/10.17705/1jais.00685

Seo, S. H., Geiskkovitch, D., Nakane, M., King, C., & Young, J. E. (2015, March 2-5). Poor thing! Would you feel sorry for a simulated robot? A comparison of empathy toward a physical and a simulated robot. *Proceedings of the 10th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, 125-132. https://ieeexplore.ieee.org/abstract/document/8520653

Serholt, S. (2018). Breakdowns in children's interactions with a robotic tutor: A longitudinal study. *Computers in Human Behavior*, *81*, 250-264. https://doi.org/10.1016/j.chb.2017.12.030

Seyama, J., & Nagayama, R. S. (2007). The uncanny valley: Effect of realism on the impression of artificial human faces. *Presence: Teleoperators and Virtual Environments, 16*(4), 337-351. https://doi.org/10.1162/pres.16.4.337

Shank, D. B., & DeSanti, A. (2018). Attributions of morality and mind to artificial intelligence after real-world moral violations. *Computers in Human Behavior*, *86*, 401-411. https://doi.org/10.1016/j.chb.2018.05.014

Shank, D. B., North, M., Arnold, C., & Gamez, P. (2021). Can mind perception explain virtuous character judgments of artificial intelligence? *Technology, Mind, and Behavior*, *2*(3), 1-12. https://doi.org/10.1037/tmb0000047

Sharma, A., Rathi, Y., Patni, V., & Sinha, D. K. (2021, June 25-27). A systematic review of assistance robots for elderly care. *Proceedings of the 2021 International Conference on Communication Information and Computing Technology (ICCICT),* 1-6. https://doi.org/10.1109/ICCICT50803.2021.9510142

Sherman, D. K., Hartson, K. A., Binning, K. R., Purdie-Vaughns, V., Garcia, J., Taborsky-Barba, S., Tomassetti, S., Nussbaum, A. D., & Cohen, G. L. (2013). Deflecting the trajectory and changing the narrative: How self-affirmation affects academic performance and motivation under identity threat. *Journal of Personality and Social Psychology, 104*(4), 591-618. https://doi.org/10.1037/a0031495

Shone, N., Ngoc, T. N., Phai, V. D., & Shi, Q. (2018). A deep learning approach to network intrusion detection. *IEEE Transactions on Emerging Topics in Computational Intelligence, 2*(1), 41-50. https://doi.org/10.1109/TETCI.2017.2772792

Silver, D., Huang, A., Maddison, C. J., Guez, A., Sifre, L., van den Driessche, G., Schrittwieser, J., Anonoglou, I., Panneershelvam, V., Lanctot, M., Dieleman, S., Grewe, D., Nham, J., Kalchbrenner, N., Sutskever, I., Lillicrap, T., Leach, M., Kavukcuoglo, K., Graepel, T., & Hassabis, D. (2016). Mastering the game of Go with deep neural networks and tree search. *Nature*, *529*(7587), 484-489. https://doi.org/10.1038/nature16961

Silver, D., Schrittwieser, J., Simonyan, K., Antonoglou, I., Huang, A., Guez, A., Hubert, T., Baker, L., Lai, M., Bolton, A., Chen, Y., Lillicrap, T., Hui, F., Sifre, L., van den Driessche, G., Graepel, T., & Hassabis, D. (2017). Mastering the game of Go without human knowledge. *Nature*, *550*(7676), 354-359. https://doi.org/10.1038/nature24270

Simonsohn, U. (2014, March 12). [17] No-way interactions. *Data Colada.* http://datacolada.org/17

Sindermann, C., Sha, P., Zhou, M., Wernicke, J., Schmitt, H. S., Li, M., Sariyska, R., Stravou, M., Becker, B., & Montag, C. (2021). Assessing the attitude towards artificial intelligence: Introduction of a short measure in German, Chinese, and English language. *KI-Künstliche Intelligenz*, *35*(1), 109-118. https://doi.org/10.1007/s13218-020-00689-0

Singer, T., & Lamm, C. (2009). The social neuroscience of empathy. *Annals of the New York Academy of Sciences*, *1156*(1), 81-96. https://doi.org/10.1111/j.1749-6632.2009.04418.x

Smith, E. R., Sherrin, S., Fraune, M. R., & Šabanović, S. (2020). Positive emotions, more than anxiety or other negative emotions, predict willingness to interact with robots. *Personality and Social Psychology Bulletin*, *46*(8), 1270-1283. https://doi.org/10.1177/0146167219900439

Smith, R. H., & Kim, S. H. (2007). Comprehending envy. *Psychological Bulletin*, *133*(1), 46-64. https://doi.org/10.1037/0033-2909.133.1.46

Spatola, N., & Wudarczyk, O. A. (2021). Ascribing emotions to robots: Explicit and implicit attribution of emotions and perceived robot anthropomorphism. *Computers in Human Behavior*, *124*, Article 106934. https://doi.org/10.1016/j.chb.2021.106934

Stapels, J. G., & Eyssel, F. (2022). Robocalypse? Yes, please! The role of robot autonomy in the development of ambivalent attitudes towards robots. *International Journal of Social Robotics*, *14,* 683-697. https://doi.org/10.1007/s12369-021-00817-2

Steain, A., Stanton, C. J., & Stevens, C. J. (2019). The black sheep effect: The case of the deviant ingroup robot. *PLoS ONE*, *14*(10), Article e0222975. https://doi.org/10.1371/journal.pone.0222975

Stein, J.-P., & Ohler, P. (2017). Venturing into the uncanny valley of mind–The influence of mind attribution on the acceptance of human-like characters in a virtual reality setting. *Cognition*, *160*, 43-50. https://doi.org/10.1016/j.cognition.2016.12.010

Stein, J.-P., & Ohler, P. (2018). Saving face in front of the computer? Culture and attributions of human likeness influence users' experience of automatic facial emotion recognition. *Frontiers in Digital Humanities*, *5,* Article 18. https://doi.org/10.3389/fdigh.2018.00018

Stein, J.-P., Appel, M., Jost, A., & Ohler, P. (2020). Matter over mind? How the acceptance of digital entities depends on their appearance, mental prowess, and the interaction between both. *International Journal of Human-Computer Studies, 142*, Article 102463. https://doi.org/10.1016/j.ijhcs.2020.102463

Stein, J.-P., Liebold, B., & Ohler, P. (2019). Stay back, clever thing! Linking situational control and human uniqueness concerns to the aversion against autonomous technology. *Computers in Human Behavior, 95,* 73-82. https://doi.org/10.1016/j.chb.2019.01.021

Stephan, W. G., Ybarra, O., & Bachman, G. (1999). Prejudice toward immigrants 1. *Journal of Applied Social Psychology, 29*(11), 2221-2237. https://doi.org/10.1111/j.1559-1816.1999.tb00107.x

Sundar, S. S. (2020). Rise of machine agency: A framework for studying the psychology of human–AI interaction (HAII). *Journal of Computer-Mediated Communication, 25*(1), 74-88. https://doi.org/10.1093/jcmc/zmz026

Swiderska, A., & Küster, D. (2020). Robots as malevolent moral agents: Harmful behavior results in dehumanization, not anthropomorphism. *Cognitive Science, 44*(7), Article e12872. https://doi.org/10.1111/cogs.12872

Takayama, L., & Pantofaru, C. (2009, October 10-15). Influences on proxemic behaviors in human-robot interaction. *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems,* 5495-5502. https://doi.org/10.1109/IROS.2009.5354145

Tamir, M. (2016). Why do people regulate their emotions? A taxonomy of motives in emotion regulation. *Personality and Social Psychology Review, 20*(3), 199-222. https://doi.org/10.1177/1088868315586325

Taylor, J., Weiss, S. M., & Marshall, P. J. (2020). “Alexa, how are you feeling today?”: Mind perception, smart speakers, and uncanniness. *Interaction Studies, 21*(3), 329-352. https://doi.org/10.1075/is.19015.tay

Tesser, A. (1988). Toward a self-evaluation maintenance model of social behavior. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (pp. 181-227). Academic Press.

Thakur, N., & Han, C. Y. (2018, November 1-3). A hierarchical model for analyzing user experiences in affect aware systems. *Proceedings of the 9th Annual Information Technology, Electronics and Mobile Communication Conference,* 783-788. https://doi.org/10.1109/IEMCON.2018.8614787

Tu, Y. C., Chien, S. E., & Yeh, S. L. (2020). Age-related differences in the uncanny valley effect. *Gerontology, 66*(4), 382-392. https://doi.org/10.1159/000507812

van den Brule, R., Dotsch, R., Bijlstra, G., Wigboldus, D. H., & Haselager, P. (2014). Do robot performance and behavioral style affect human trust? *International Journal of Social Robotics, 6*(4), 519-531. https://doi.org/10.1007/s12369-014-0231-5

van Esch, P., Black, J. S., & Ferolie, J. (2019). Marketing AI recruitment: The next phase in job application and selection. *Computers in Human Behavior, 90,* 215-222. https://doi.org/10.1016/j.chb.2018.09.009

Vanman, E. J., & Kappas, A. (2019). “Danger, Will Robinson!” The challenges of social robots for intergroup relations. *Social and Personality Psychology Compass, 13*(8), Article e12489. https://doi.org/10.1111/spc3.12489

Vidler, A. (1992). *The architectural uncanny: Essays in the modern unhomely.* The MIT Press.

Waldrop, M. M. (2016). The chips are down for Moore's law. *Nature, 530,* 144-147. https://doi.org/10.1038/530144a

Walsh, C. G., Ribeiro, J. D., & Franklin, J. C. (2018). Predicting suicide attempts in adolescents with longitudinal clinical data and machine learning. *Journal of Child Psychology and Psychiatry, 59*(12), 1261-1270. https://doi.org/10.1111/jcpp.12916

Wang, S., Lilienfeld, S. O., & Rochat, P. (2015). The uncanny valley: Existence and explanations. *Review of General Psychology, 19*(4), 393-407. https://doi.org/10.1037/gpr0000056

Ward, A. F., Olsen, A. S., & Wegner, D. M. (2013). The harm-made mind: Observing victimization augments attribution of minds to vegetative patients, robots, and the dead. *Psychological Science, 24*(8), 1437-1445. https://doi.org/10.1177/0956797612472343

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

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. http://dx.doi.org/10.1016/j.jesp.2014.01.005

Waytz, A., Klein, N., & Epley, N. (2013). Imagining other minds: Anthropomorphism is hair-triggered but not hare-brained. In M. Taylor (Ed.), *The Oxford handbook of the development of imagination* (pp. 272–287). Oxford University Press.

Webb, T. L., & Sheeran, P. (2006). Does changing behavioral intentions engender behavior change? A meta-analysis of the experimental evidence. *Psychological Bulletin, 132*(2), 249-268. https://doi.org/10.1037/0033-2909.132.2.249

Wegner, D. M., & Gray, K. (2016). *The mind club: Who thinks, what feels, and why it matters.* Viking.

Weis, P. P., & Wiese, E. (2017). Cognitive conflict as possible origin of the uncanny valley. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 61*(1), 1599-1603. https://doi.org/10.1177/1541931213601763

Weisman, K., Dweck, C. S., & Markman, E. M. (2017). Rethinking people's conceptions of mental life. *Proceedings of the National Academy of Sciences, 114*(43), 11374-11379. https://doi.org/10.1073/pnas.1704347114

Weiss, A., Igelsböck, J., Wurhofer, D., & Tscheligi, M. (2011). Looking forward to a "robotic society"? *International Journal of Social Robotics, 3*(2), 111-123. https://doi.org/10.1007/s12369-010-0076-5

Wiese, E., Weis, P. P., Bigman, Y., Kapsaskis, K., & Gray, K. (2022). It's a match: Task assignment in human–robot collaboration depends on mind perception. *International Journal of Social Robotics, 14*(1), 141-148. https://doi.org/10.1007/s12369-021-00771-z

Wiggins, J. S. (1979). A psychological taxonomy of trait-descriptive terms: The interpersonal domain. *Journal of Personality and Social Psychology, 37*(3), 395-412. https://doi.org/10.1037/0022-3514.37.3.395

Witteman, C., van den Bercken, J., Claes, L., & Godoy, A. (2009). Assessing rational and intuitive thinking styles. *European Journal of Psychological Assessment, 25*(1), 39-47. https://doi.org/10.1027/1015-5759.25.1.39

Wojciszke, B. (2005). Morality and competence in person-and self-perception. *European Review of Social Psychology, 16*(1), 155-188. https://doi.org/10.1080/10463280500229619

Wojciszke, B., & Białobrzeska, O. (2014). Agency versus communion as predictors of self-esteem: Searching for the role of culture and self-construal. *Polish Psychological Bulletin, 45*(4), 469-479. https://doi.org/0.2478/ppb-2014-0057

Woo, W. L. (2020). Future trends in I&M: Human-machine co-creation in the rise of AI. *IEEE Instrumentation & Measurement Magazine, 23*(2), 71-73. https.//doi.org/10.1109/MIM.2020.9062691

Wu, Y. H., Fassert, C., & Rigaud, A. S. (2012). Designing robots for the elderly: Appearance issue and beyond. *Archives of Gerontology and Geriatrics, 54*(1), 121-126. https://doi.org/10.1016/j.archger.2011.02.003

Xu, M., David, J. M., & Kim, S. H. (2018). The fourth industrial revolution: Opportunities and challenges. *International Journal of Financial Research, 9*(2), 90-95. https://doi.org/10.5430/ijfr.v9n2p9

Yam, K. C., Bigman, Y. E., Tang, P. M., Ilies, R., De Cremer, D., Soh, H., & Gray, K. (2020). Robots at work: People prefer—and forgive—service robots with perceived feelings. *Journal of Applied Psychology, 106*(10), 1557-1572. http://dx.doi.org/10.1037/apl0000834

Yam, K. C., Goh, E. Y., Fehr, R., Lee, R., Soh, H., & Gray, K. (2022). When your boss is a robot: Workers are more spiteful to robot supervisors that seem more human. *Journal of Experimental Social Psychology, 102*, Article 104360. https://doi.org/10.1016/j.jesp.2022.104360

Yang, G. Z. J., Nelson, B., Murphy, R. R., Choset, H., Christensen, H., Collins, S. H., Dario, P., Goldberg, K., Ikuta, K., Jacobstein, N., Kragic, A., Taylor, R. H., & McNutt, M. (2020). Combating COVID-19—The role of robotics in managing public health and infectious diseases. *Science Robotics*, *5*(40), Article eabb5589. https://doi.org/10.1126/scirobotics.abb5589

Ybarra, O., Chan, E., Park, H., Burnstein, E., Monin, B., & Stanik, C. (2008). Life's recurring challenges and the fundamental dimensions: An integration and its implications for cultural differences and similarities. *European Journal of Social Psychology, 38*(7), 1083-1092. https://doi.org/10.1002/ejsp.559

Yee, D. M., Adams, S., Beck, A., & Braver, T. S. (2019). Age-related differences in motivational integration and cognitive control. *Cognitive, Affective, & Behavioral Neuroscience, 19*(3), 692-714. https://doi.org/10.3758/s13415-019-00713-3

Yin, J., Wang, S., Guo, W., & Shao, M. (2021). More than appearance: The uncanny valley effect changes with a robot's mental capacity. *Current Psychology*. Advance online publication. https://doi.org/10.1007/s12144-021-02298-y

Yogeeswaran, K., Złotowski, 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, March 5-8). Human–robot similarity and willingness to work with a robotic co-worker. *Proceedings of the 13th ACM/IEEE International Conference on Human-Robot Interaction (HRI),* 251-260. https://doi.org/10.1145/3171221.3171281

Young, A. D., & Monroe, A. E. (2019). Autonomous morals: Inferences of mind predict acceptance of AI behavior in sacrificial moral dilemmas. *Journal of Experimental Social Psychology, 85,* Article 103870. https://doi.org/10.1016/j.jesp.2019.103870

Young, A., Khalil, K. A., & Wharton, J. (2018). Empathy for animals: A review of the existing literature. *Curator: The Museum Journal, 61*(2), 327-343. https://doi.org/10.1111/cura.12257

Zafari, S., & Koeszegi, S. T. (2020). Attitudes toward attributed agency: Role of perceived control. *International Journal of Social Robotics, 13*(8), 2071-2080. https://doi.org/10.1007/s12369-020-00672-7

Zeng, L., Li, L., & Duan, L. (2012). Business intelligence in enterprise computing environment. *Information Technology and Management, 13*(4), 297-310. https://doi.org/10.1007/s10799-012-0123-z

Zhang, C., & Lu, Y. (2021). Study on artificial intelligence: The state of the art and future prospects. *Journal of Industrial Information Integration, 23,* Article 100224. https://doi.org/10.1016/j.jii.2021.100224

Zhang, G., Zhong, J., & Ozer, M. (2020). Status threat and ethical leadership: A power-dependence perspective. *Journal of Business Ethics, 161*(3), 665-685. https://doi.org/10.1007/s10551-018-3972-5

Zhang, T., Zhang, W., Qi, L., & Zhang, L. (2016, August 1-3). Falling detection of lonely elderly people based on NAO humanoid robot. *Proceedings of the 2016 IEEE International Conference on Information and Automation (ICIA),* 31-36. https://doi.org/10.1109/ICInfA.2016.7831793

Złotowski, J., Proudfoot, D., Yogeeswaran, K., & Bartneck, C. (2015). Anthropomorphism: Opportunities and challenges in human-robot interaction. *International Journal of Social Robotics, 7*(3), 347-360. https://doi.org/10.1007/s12369-014-0267-6

Złotowski, J., Yogeeswaran, K., & Bartneck, C. (2017). Can we control it? Autonomous robots threaten human identity, uniqueness, safety, and resources. *International Journal of Human-Computer Studies, 100,* 48-54. https://doi.org/10.1016/j.ijhcs.2016.12.008

Appendix

Appendix A

Project 1 — Experiment 1: Concerns About Human Identity Scale

Composition of five items:

Three Items by Stein et al. (2019)

- This robot does not change the value of humanity at all.
- This robot would mess with the order of the world.
- This robot would reduce the significance of us humans.

Two Items by Kamide et al. (2012)

- This robot seems to reduce the meaning of a person’s existence.
- This robot seems to break the relationship between humans.

Appendix B

Project 3 — Experiment 1: Screenplay

Herzlich willkommen zur Studie!

Bitte wenden Sie sich zunächst dem Fragebogen zu und lesen die Einverständniserklärung durch. Sobald Sie Ihr Einverständnis gegeben haben und im Fragebogen dazu aufgefordert werden, geben Sie der Versuchsleitung bitte ein Zeichen.

[Versuchsperson liest Einverständniserklärung]

Vielen Dank. In den nächsten Minuten werden Sie gemeinsam mit Roboter NAO verbale Zuordnungsaufgaben bearbeiten und Wörter auswählen, die Ihrer Meinung nach am ehesten gut zusammenpassen.

[NAO: „Hallo! Ich freue mich, gemeinsam mit dir an den Aufgaben zu arbeiten.“]

Da Sie beide hier in Würzburg am Institut beschäftigt sind und sich somit für Kommunikation und Technik interessieren, möchten wir herausfinden, wie Mensch und Roboter gemeinsam in einer verbalen Aufgabe abschneiden. Solche verbale Aufgaben werden genutzt, um analytisches Denken zu erfassen, einer wichtigen Fähigkeit in vielen Bereichen (akademische Werdegänge, Berufsleben, etc.).

Wenn Sie sich gleich wieder dem Fragebogen zuwenden, werden Sie und NAO jeweils acht Aufgaben in fünf Durchgängen präsentiert bekommen. Es gibt also 40 Aufgaben. Manchmal werden Sie den Eindruck haben, dass es keine eindeutige Lösung gibt, dann wählen Sie das Wort aus, das Sie am ehesten für geeignet halten und lassen Ihrer Kreativität freien Lauf. Es ist wichtig, dass (wie bei jeder Aufgabe, die man im Team löst) jede*r so viel zur Lösung der Aufgaben beiträgt, wie er/sie kann – Mensch und Maschine. NAO kann auf ein neuronales Netz (genau genommen ein rekurrentes neuronales Netz) zur Worterkennung zugreifen und ist per WLAN mit der Datenbank verbunden, in die auch Ihre Antworten auf die Aufgaben eingespeist werden.

Nach jeder Runde bekommen Sie ein Feedback über Ihre Leistung aufbauend auf den Einträgen in der Datenbank. Dabei wird verglichen, ob Sie oder NAO in der gleichen Zeit mehr zu der Lösung der Aufgaben beigetragen haben. Der Beitrag der Teilnehmenden zur Aufgabe richtet sich nach der individuellen Anzahl der richtig gelösten Aufgaben in Kombination mit der Länge, Kreativitätsanforderung und Schwierigkeit dieser Aufgaben. Es besteht die Möglichkeit, dass Sie oder NAO bei sehr gutem Abscheiden bei den Aufgaben bevorzugt bei einer Verlosung eines Gutscheins für Strom in Höhe von 25 Euro behandelt werden. Es lohnt sich also, sich anzustrengen! Gibt es dazu Fragen?

[NAO: „Nein, ich habe keine Fragen.“]

Dann bitte ich Sie nun, sich wieder dem Fragebogen zuzuwenden. Ihre Eingaben werden automatisch an die Datenbank übermittelt, in die auch Naos Antworten eingespeist werden. Aufbauend auf den Ergebnissen der Auswertung wird Ihnen Ihr Feedback in der Fragebogenoberfläche angezeigt.

[NAO: „Los geht’s!“]

[Versuchsperson bearbeitet Aufgaben mit eingespielten manipulierten Feedback (siehe Supplement von Projekt 3). NAO bearbeitet Aufgaben im autonomous life scheinbar ebenfalls. Nach den fünf Runden und dem finalen Feedback:

- Falls NAO mehr zur Lösung der Aufgaben beiträgt: NAO: „Super, ich mag es wenn ich in solchen Aufgaben richtig gut abschneide.“
- Falls NAO weniger zur Lösung der Aufgaben beiträgt: NAO: „Schade, da hast du deutlich mehr Wissen einbringen können als ich.“]

Bitte wenden Sie sich nun den abschließenden Fragen zu.

[Abschließender Fragebogen]

Haben Sie noch Fragen zu den Inhalten der Studie?

Möchten Sie Ihre E-Mail-Adresse mitteilen, sodass wir Sie bei der Verlosung von 25 Euro berücksichtigen können? Die Daten werden getrennt von den Fragebogendaten abgelegt. Die Versuchspersonenstunden werden selbstverständlich auch zügig verbucht.

Wir bedanken uns sehr für Ihre Teilnahme und wünschen einen angenehmen Tag!

[NAO: „Tschüss, hab noch einen schönen Tag!“]

Appendix C

Project 3: German Items Used in the Questionnaires

Pettit et al. (2013): Status Threat (Experiment 1: Robot, Experiment 2: Artificial Intelligence)

- Bald wird ein Roboter/eine künstliche Intelligenz einen höheren Status am Arbeitsplatz haben als ich.
- Bald wird ein Roboter/eine künstliche Intelligenz ein höheres Ansehen am Arbeitsplatz haben als ich.
- Bald wird ein Roboter/eine künstliche Intelligenz einen höheren Bekanntheitsgrad am Arbeitsplatz haben als ich.
- Bald wird ein Roboter/eine künstliche Intelligenz mehr Bewunderung am Arbeitsplatz haben als ich.

Robinson et al. (2018): Willingness to Interact (Experiment 1: Robot, Experiment 2: Artificial Intelligence)

- Ich würde mit dem Roboter/der künstlichen Intelligenz interagieren.
- Ich würde den Roboter/die künstlichen Intelligenz um Rat fragen.
- Ich würde den Roboter/die künstlichen Intelligenz bitten, mir regelmäßig bei einer Aufgabe zu helfen.
- Ich würde über einen längeren Zeitraum mit dem Roboter/der künstlichen Intelligenz interagieren wollen.
- Ich würde Zeit mit dem Roboter verbringen (only Experiment 1).

Davis (1989): Perceived Usefulness (only Experiment 2)

- Eine solche künstliche Intelligenz bei zukünftigen Aufgaben zu nutzen würde mich diese schneller erledigen lassen.
- Eine solche künstliche Intelligenz bei zukünftigen Aufgaben zu nutzen würde meine Arbeitsleistung verbessern.

- Eine solche künstliche Intelligenz bei zukünftigen Aufgaben zu nutzen würde meine Produktivität erhöhen.
- Eine solche künstliche Intelligenz bei zukünftigen Aufgaben zu nutzen würde meine Effektivität bei meinen Aufgaben erhöhen.
- Eine solche künstliche Intelligenz bei zukünftigen Aufgaben zu nutzen würde es mir leichter machen, meine Aufgaben zu erfüllen.
- Ich würde den Einsatz von künstlicher Intelligenz für meine Aufgaben nützlich finden.

Dang and Liu (2022a): Mindset about Human Minds (only Experiment 2)

- Der Verstand eines Menschen ist etwas sehr Grundlegendes und kann nicht viel verändert werden.
- Ob eine Person verständig oder unverständig ist, kann nicht wirklich verändert werden.
- Menschen können nicht wirklich ändern, wie viel Verstand sie haben. Manche Menschen sind sehr verständig und manche nicht, und daran kann man nicht viel ändern.
- Unabhängig davon, wer jemand ist, kann ein Mensch immer das Level seines Verstandes ändern.
- Die Menschen können immer ändern, wie viel Verstand sie im Allgemeinen haben.
- Jeder kann ändern, wie viel Verstand ein Mensch hat.

Sindermann et al. (2021): Attitude towards AI (only Experiment 2, Covariate)

- Ich habe Angst vor künstlicher Intelligenz.
- Ich vertraue künstlicher Intelligenz.
- Künstliche Intelligenz wird die Menschheit zerstören.
- Künstliche Intelligenz wird eine Bereicherung für die Menschheit sein.

- Künstliche Intelligenz wird für viel Arbeitslosigkeit sorgen.