



Humanoid Robots – Artificial. Human-like. Credible? Empirical Comparisons of Source Credibility Attributions Between Humans, Humanoid Robots, and Non-human-like Devices

Marcel Finkel¹ · Nicole C. Krämer¹

Accepted: 23 March 2022 / Published online: 23 April 2022
© The Author(s) 2022

Abstract

Source credibility is known as an important prerequisite to ensure effective communication (Pornpitakpan, 2004). Nowadays not only humans but also technological devices such as humanoid robots can communicate with people and can likewise be rated credible or not as reported by Fogg and Tseng (1999). While research related to the machine heuristic suggests that machines are rated more credible than humans (Sundar, 2008), an opposite effect in favor of humans' information is supposed to occur when algorithmically produced information is wrong (Dietvorst, Simmons, and Massey, 2015). However, humanoid robots may be attributed more in line with humans because of their anthropomorphically embodied exterior compared to non-human-like technological devices. To examine these differences in credibility attributions a 3 (source-type) x 2 (information's correctness) online experiment was conducted in which 338 participants were asked to either rate a human's, humanoid robot's, or non-human-like device's credibility based on either correct or false communicated information. This between-subjects approach revealed that humans were rated more credible than social robots and smart speakers in terms of trustworthiness and goodwill. Additionally, results show that people's attributions of theory of mind abilities were lower for robots and smart speakers on the one side and higher for humans on the other side and in part influence the attribution of credibility next to people's reliance on technology, attributed anthropomorphism, and morality. Furthermore, no main or moderation effect of the information's correctness was found. In sum, these insights offer hints for a human superiority effect and present relevant insights into the process of attributing credibility to humanoid robots.

Keywords Human-robot interaction · Source credibility · Source type · Embodiment · Theory of mind

1 Introduction

A look at the proposed application contexts of social robots shows that these technological devices can be used in a wide variety of scenarios in the long term, including as helpers in elderly care, museum guides, or even teachers [1]. One similarity between all of these use cases is that they rely to a greater or lesser extent on communication between robots and humans and therefore on the exchange of information. Many social robots are even shaped like humans (humanoids) signaling that users should exhibit routinized social communication behaviors learned from interactions with humans. However, social robots must not only be able to engage in

conversations but also need to be perceived as competent, trustworthy, and caring in order to explain, report, and teach successfully. In other words, effective human-robot interaction can be supported by social robots that are perceived as credible sources of information. Because in the long run social robots are expected to fulfill complex social tasks and roles due to increased abilities, the role of information sharing and simultaneously the importance of source credibility will steadily increase. With regard to research on interpersonal communication it has been shown that a credible source's information can have a strong impact on people's opinions and behavioral intentions and as a consequence supports trust-building [2]. In addition, credibility has been shown to be important not only for human-related attributions but also for technology artifacts including websites [3] and technological devices, such as smart speakers [4] and social robots [5]. Some research works have already investigated how people make credibility evaluations related to these

✉ Marcel Finkel
marcel.finkel@uni-due.de

¹ University of Duisburg-Essen, Forsthausweg 2, 47057
Duisburg, Germany

technological artifacts. One example refers to the machine heuristic by Sundar [6] who postulated that machines (or rather devices) are rated more credible due to perceived impartiality. On the contrary, other research works suggest that people seem to reject input from an algorithmic source if it showed errors in the past [7]. In these cases, it was demonstrated that reliance on the information given by humans was higher even when people witnessed a human giving incorrect information as well. Such different responses to humans on the one hand and artificial sources, on the other hand, become more complex by considering social humanoid robots that are able to blur boundaries between distinct categories of humans and technological devices. Because of their anthropomorphic embodiment and social demand characteristics people experience humanoid robots differently than non-human-like devices such as televisions or laptops and more often show behaviors derived from human interactions [8]. To a greater or lesser degree, therefore, these robots may be able to negate aversions to inaccurate information from artificial sources, thus promoting long-term interactions and ensuring reliable information-sharing processes.

The main question we want to address with this paper, therefore, concentrates on the extent to which a source's type (human, humanoid robot, non-human-like device) causes different credibility attributions depending on the correctness of the communicated information. Furthermore, we want to investigate the degree to which human-related constructs like people's perception of a robot's theory of mind abilities are attributed to humanoid robots and contribute to a more beneficial credibility rating.

2 Theoretical Background

2.1 The Concept of Credibility

Due to the huge importance and quantity of information sharing in people's daily lives, credibility is of great importance in a variety of contexts where people act as senders of relevant information, such as testimonial advertisements [9], politicians on social media [10], or witnesses to crime scenes [11]. This sender-related credibility can be described as the "believability of a source" (p. 211) [3]. In other words, it represents an individually perceived attribute ascribed to a specific information-delivering entity, rating the probability of it being a competent, trustworthy, and good willing source. Competence refers to the ability to have access to topic-relevant knowledge as well as expertise while trustworthiness includes beliefs about a source's honesty and independence from persuasion or manipulation attempts [12]. The last sub-construct called goodwill can be defined as the perceived interest of a source for one's own personal well-being [13].

In combination, these three are the constituents of a source's credibility.

However, the concept of credibility attributions is not exclusively applicable to human sources. In 1999, Fogg and Tseng [14] published a conclusive differentiation of credibility constructs referring to technological devices such as computers. The authors distinguished the concepts of presumed, reputed, surface, and experienced credibility. The first two constructs describe credibility perceptions that are formed before getting into the first interaction with a respective device. In contrast, the last two constructs define credibility evaluations based on superficial first impressions (surface credibility) as well as on a more profound basis after some interaction took place (experienced credibility). However, even if people are not depending on reputed or presumed judgments of credibility but can rather make credibility judgments based on their own experiences, these evaluations will not be an objective measurement but rather a person's estimation attempt based on available information, experiences, and attitudes. Therefore, credibility has to be regarded as the result of an information receiver's attribution process and not as a source's stable characteristic [14]. This individual attribution process has the potential to cause different ratings of credibility which explains why not all people evaluate technology devices as similar credible and why consequently incredulity or gullibility errors can occur.

Because the number of technological innovations that are able to serve as sources in information sharing processes is increasing steadily (e.g., voice assistants, social robots, chatbots, etc.) the relevance to investigate the process of forming credibility attributions to such artificial sources becomes apparent. Research on human-computer interaction has already recognized this need and analyzed credibility attributions to computers and social humanoid robots and supported the theoretical assumptions postulated by Fogg and Tseng [14]. The empirical results of these studies will be discussed after summarizing why in the first place, people are able to make credibility attributions about inanimate technological devices in a similar way as they evaluate human beings.

2.2 Media Equation and Anthropomorphism

People often interact socially with technological devices in ways that are highly influenced by their knowledge about interactions with other humans. Reeves and Nass [15] analyzed this tendency towards social behavior empirically and stated that (related to a specific computer product) people "when in doubt, treat it as human" (p. 22) [15]. This demonstrates their point of view that people make use of their social conventions even when confronted with a situation in which no demand for such social behavior is expressed, because technological devices do not have an internal need

for sociality. Reeves and Nass [15] thoroughly explained and argued that this behavior is not necessarily a product of conscious evaluations but rather an unaware and automatically triggered process, called media equation. Other researchers integrated this phenomenon into a dual model of anthropomorphism, which on the one hand explains that media equation is related to a fast, less effort needing, and intuitive type of attribution process and regard it as an outcome of implicit anthropomorphism [16]. The most important factor to let people elicit unaware social reactions to objects like computers, smart speakers, or other technological devices like robots is the presence of a sufficient number of social cues as they can activate people's internalized communication scripts equally to human-human interactions. These cues that trigger the process of media equation, include the usage of human language in text or speech, signs for the fulfillment of common social roles (e.g., a tutor or a colleague) as well as a reciprocal interaction component by reacting to others' behaviors [17]. Thus, one way to show these cues is the usage of verbal communication, a feature inherent in many technological devices. Empirical research findings demonstrate that the phenomenon of media equation is not only relevant with regard to computers but also for interactions with other technological entities such as virtual agents [18], social robots [19], and smart speakers [20]. It's because these devices and applications can make use of speech systems supported by artificial intelligence, that they are able to use language, interactivity, and social roles to successfully trigger the media equation.

On the other hand, this implicit and mindless form of social reactions can be distinguished from explicit anthropomorphism which requires more mental capacities, is slower, and includes an apparent, conscious willingness to provide human-like treatment to objects such as technological devices [16]. In contrast to devices such as smart speakers, robots are often physically and anthropomorphically embodied to mimic the appearance of humans in order to support this intuitive expression of even more social behaviors [1]. In sum, these human-like physical characteristics such as a head, a mouth, eyes, limbs, and a torso can be defined as a robot's anthropomorphic design [21]. Users' perceptions of an anthropomorphic design seem to be especially influenced by four major robot characteristics identified during a principal components' analysis by Phillips, Zhao, Ullman, and Malle [22]: a robot's surface look including the existence of hair or skin, its facial features like eyes or mouth, its body manipulators (e.g., extremities, torso), and its mechanic locomotion which refers to the way a robot is able to move around in physical space. Based on these extracted dimensions, the authors developed a human-likeness score to measure the degree to which a robot implements an anthropomorphic design. Their measure of human-likeness is a continuous score ranging from zero ("Not human-like at all")

to one hundred ("Just like a human") (p.110) [22]. It shows that human-likeness cannot be regarded as a dichotomous variable (meaning a robot can only have an anthropomorphic design or not) but rather as a continuous property that can be expressed and measured in continually increasing degrees. At best, this anthropomorphic design can ease interactions for users because they consciously react socially to these higher demand characteristics. One example of this interaction's facilitation is the exploitation of the cuteness expressing Baby Schema [23]. Here, a robot's facial features are constructed in ways to replicate the key characteristics of human infants' faces (e.g., big head, big eyes, high forehead, etc.). This implementation of cuteness helps to support an immediate understanding of the intended role constellation between a user as the protector and the robot as an object in need of protection and to promote strong caring reactions for the as vulnerable perceived device. Although a robot's anthropomorphic design by itself is no guarantee for such explicit anthropomorphism to work because user characteristics [24] and context-specific environment variables influence this process as well [21], the design can help to shape expectancies about how the interaction with the robot will look like. However, it is also important to mention that the variety of robots' anthropomorphic designs is huge and therefore even robots that are somewhat similar to each other can still be evaluated differently [25].

In sum, although the use of language is an important tool to allow the media equation to work, the most often used key feature to distinguish social robots from other devices is their anthropomorphically designed embodiment, which is an integral component of several robot taxonomies [1,26]. If, in addition, it is also taken into account that humanoid robots are very frequently used in practice, the question arises whether a social robot's anthropomorphic embodiment, including its human-like components, has a particular influence on people's evaluations about it. More precisely put, does the human-likeness explain differences when it comes to attributing credibility in comparison to human and non-human-like devices?

2.3 Credible, Social, Humanoid Robots

Based on the assumptions from Fogg and Tseng [14] it can be deduced and shown that people are able to make credibility attributions about robots. As a result, this attribution process has been a topic of interest for many years. Researchers examined that a robot's credibility can be supported for example by improving its rhetorical abilities [27], reducing suspicion [28], or using non-verbal cues [29]. Such papers deliver meaningful insights into the process of credibility attribution related to robots. However, they do not allow for comparisons with humans or technological devices without anthropomorphic embodiment. As already explained, humanoid robots are

meant to work as assistants in elderly care, museum guides, or as teachers. Because these are jobs that can only be taken on by humans nowadays, it seems highly relevant to investigate if credibility differences or similarities to humans, on the one hand, are recognizable and how non-human-like technological devices' performances are evaluated on the other hand. One line of reasoning for how technological devices in contrast to human sources influence perceptions of credibility is the machine-heuristic [6]. It describes the idea "that if a machine chose the story, then it must be objective in its selection and free from ideological bias" (p. 83) [6] because people make use of their stereotypes related to machines (or rather devices) which most likely include associations of objectivity rather than partiality. In other words, people assume a device to work on a mere functional basis without having an own agenda and thus expect a lesser threat of manipulation which should result in better ratings of trustworthiness and goodwill [12]. A similar phenomenon that includes positive expectancies towards artificial entities is the concept of algorithm appreciation, postulating that people are more in favor of recommendations made by algorithms compared to human suggestions. Logg, Minson, and Moore [30] investigated the robustness of this effect for several application contexts provided that the algorithmic information was correct. However, these studies did not focus on robots or specific devices and thus do not allow to draw generalized conclusions independent from device-specific features such as voice or embodiment. While vocal cues already showed to be able to trigger the media equation [17] the use of anthropomorphic embodiment influences people's understanding of the interaction more consciously. Since social robots are able to make use of both features, voice and anthropomorphic embodiment, we expect that this more profound resemblance to humans will also cause more similar credibility attributions between them and humans. On the other hand, devices such as smart speakers which do only make use of human-like voices should receive less similar attributions compared to human sources. Provided that the communicated information is correct, we thus propose the following hypothesis:

Hypothesis H1 Correct communicated information leads to a higher rating of a source's credibility in cases of non-human-like devices compared to human sources and anthropomorphically embodied robots.

In addition to the concept of algorithm appreciation, Dietvorst, Simmons, and Massey [7] collected findings supporting the assumption of a phenomenon's existence called algorithm aversion. Here, a shift in preference from an algorithmic information source to a human source occurs when the algorithmically given information was identified as incorrect. At first glance, algorithm appreciation seems opposed to the concept of algorithm aversion, but while algorithm appreciation refers to people's presumed superiority of algorithms,

algorithm aversion occurs when people are witnessing an algorithm failing at giving a correct output [7,31]. In such cases, the algorithm seems to no longer be the preferred source of information, but instead, human suggestions are accepted at a better rate even if the respective human source is giving insufficient recommendations as well. Although these findings are focused on the domain of information forecasting, the explicitly higher adverse reactions to algorithmic information output seem applicable to other domains too. When combining these findings with the often anthropomorphically embodied social robots one could argue that these robots might be able to negate these adverse effects because robots are often treated similarly to humans rather than non-human-like technological devices. That means, in cases of false information their resemblance to humans should counter the consequences of algorithm aversion and thus cause higher credibility attributions for robots and humans in comparison to non-human-like devices.

Hypothesis H2 False communicated information leads to a worse rating of a source's credibility in cases of non-human-like devices compared to human sources and anthropomorphically embodied robots.

2.4 A Robot's Theory of Mind?

As described earlier, the machine heuristic [6] is based on the idea that people do not attribute a mental state to a device that prevents it from being perceived as suspicious, falsifying, framing, or manipulating information. Although this explanation seems reasonable to some extent, other researchers have conducted experiments that give hints to suggest that people are indeed able and willing to imagine devices and especially robots to have a consciousness [32–34]. With regard to human interactions, people are able to understand that other people are intentional agents whose behaviors are not randomly chosen but linked to and depending on their inner states, beliefs, and desires. This phenomenon can be described as having a theory of mind (ToM). ToM is an ability which serves as a prerequisite for successful communication because it allows people to gain an understanding of other people's mind and intentions [35]. Although this process is primarily important in human interactions, people are able to attribute this construct to technological devices such as robots as well. Among others, Krach and colleagues [36] investigated the effect of ToM processes in human-robot interactions. Their results showed a relation between the number of human-like features and the degree to which people are attributing a mind to an artificial entity. Similar findings were obtained by Manzi et al. who revealed a higher tendency for children attributing mind to a human-like robot compared to a mechanical one [37]. Furthermore, Benninghoff and colleagues [32] investigated that people

are able to imagine humanoid robots to have a ToM and showed that this ToM attribution can for example influence people's perception of a robot's social attractiveness. Additionally, analyses of neural activation patterns showed partly similar results for human and robotic stimuli with regard to empathy [38], indicating that people's mental representations of humanoid robots might be comparable to those of humans to some extent. But also, with regard to source credibility assessments, this attribution of a ToM could be an important influencing factor. This is because credibility is not only defined by competence but also by expressing the more socially focused attributes goodwill and trustworthiness. Especially the subconstruct goodwill implicitly requires that the source has the mental abilities to know about what other people expect and desire. Thus, it seems reasonable to suggest that ToM attributions act as a foundation that supports the attribution of credibility. Due to humanoid robots' resemblance to humans, they should on the one hand receive more similar attributions compared to humans [25] and as a result on the other hand higher ones in contrast to devices without anthropomorphically embodied physiques [37].

Hypothesis H3 People will attribute a more expressed theory of mind to an anthropomorphically embodied robot than to a non-human-like device (smart speaker) but in both cases less than compared to a human source.

As explained before, it is known that it is a foundation of human communication to attribute a ToM to other human beings. Nevertheless, with regard to human-robot interactions, this attribution of inner states and intentions seems to be in conflict with the machine heuristic' assumptions [6] which postulate that devices are regarded as more credible because they do not have motives or desires to manipulate their communicated information. A possible solution to this contradiction could be to assume that attributions of mental awareness are only beneficial for credibility judgments (especially related to goodwill) in cases of correct information. Since in such situations, there are no apparent reasons to doubt a device's intentions, attributing ToM abilities to the device could have an exclusively credibility-enhancing effect, because of the device's presumably better understanding of its communication partners. On the other hand, this relationship would be reversed in cases of incorrect information. Here, attributions of ToM abilities could cause the source to be judged as manipulative rather than benevolent, with negative implications for credibility ratings. Based on these considerations, the following moderation by the communicated information's correctness is assumed.

Hypothesis H4 People's degree to which they attribute a theory of mind to a technological device affects its attributed credibility depending on the information's correctness.

2.5 Further Influencing Factors

As already explained, attributions of credibility are the results of an individual's evaluation process and therefore are highly subjective and not an inherent source characteristic [14]. For this reason, it seems important to examine further influences on credibility evaluations including user-related variables but also additional source-related attributions.

2.5.1 Machine Heuristic

The machine heuristic has been defined as a mental shortcut that lets users attribute higher credibility to devices and the information they offer [6]. This means that when seeing a robot or smart speaker the machine heuristic should be triggered and influence credibility attributions related to the source of information but also to the information's credibility as well. However, a study by Waddell [39] could not confirm a significant effect of this machines heuristic's strength on an article's rated credibility. Because we are interested in the effect of an information's correctness on source credibility attributions, we again want to investigate if the machine heuristic's strength mitigates the perception of the information's correctness in the first place, since the information's credibility is closely linked to the perception of it being seen as correct or false. Thus, the machine heuristic's influence on the information's perceived correctness will be investigated in addition to the main analyses of technological devices' credibility attributions.

Research Question 1: Does the machine heuristic influences the perception of a device's informational correctness?

2.5.2 Reliance on Technology

Due to the immense degree of digitization, people are able to rely on a large number of technological devices to provide them with information. The extent of general reliance on technological recommendations can be defined as a personal trait, referred to as reliance on technology [40] which seems relevant to be considered when examining the evaluation of single devices. Previous research already analyzed how different attitudes towards technology influence the way we approach single products (for instance technological affinity being related to less anxiety towards robots [41]). Reliance on technology is a product of habituated use of technology, such as accessing news daily via the same websites because their information proved to be the most current and accurate [42]. Thus, similar to trust in human interactions, reliance on technology could work like an underlying filter through which technological devices such as robots or smart speakers will be perceived and evaluated. It seems plausible to expect a positive relationship between reliance on technology and

credibility attributions since a higher general willingness for accepting and using technological outputs will also affect the evaluation related to the output of a specific device. However, because we lack empirical evidence of this relation, we want to analyze it (in combination with other potential credibility predictors) in our second research question.

2.5.3 Moral Agency

Assuming robots are able to give the impression of being conscious or even having a theory of mind, the interesting question arises whether humans will grant them moral decision-making abilities. Jackson and Williams argue that people do expect social robots to behave according to moral norms of humans because people automatically expand this concept of moral agency from humans to robots, thus seeing the robot as a moral agent [43]. This degree of attributed moral agency might be a useful predictor for robotic credibility attributions because it builds upon the previous assumptions of theory of mind but furthermore includes the idea of internal guidelines which assure that the robot's messages are communicated based on an ethical awareness [34]. Thus, more profound moral agency perceptions could be connected to higher credibility attributions for robots, as people will rate their behavior as more integer and bound to the users' moral norm perceptions. Again we lack empirical evidence to support our assumptions. Therefore, we intend to exploratory analyze the effect of moral agency attribution on credibility in combination with reliance on technology.

Research Question 2: What influence do other related constructs such as reliance on technology and moral agency attributions exert on a humanoid robot's credibility?

3 Method

3.1 Design

An online experiment was conducted to investigate the hypothesized influences on credibility attributions towards humanoid social robots, and to compare it to humans as well as technological devices with non-human-like embodiments based on either correct or false presented information. Therefore, the experiment employed a three (source type: human vs. humanoid robot vs. non-human-like device) x two (information's correctness: true vs. false) between-subjects design. The study's main stimulus was a short video (~60s) which included the manipulation of both independent variables. In this video, a human, a robot, or a smart speaker communicated either correct or false information during a speech about a German education reform from 2019. Thus, six different conditions were created. Before starting the study in Septem-

ber 2020, the study was preregistered on OSF.io and an approval of the university's ethical committee was obtained. Data collection ended in December 2020, after ensuring that at least fifty valid data files per condition have been gathered.

3.2 Independent Variables

3.2.1 Information's Correctness

The first independent variable, the information's correctness, was manipulated by presenting a speech with either correct or false information during the stimulus video. The speech with correct information included statements about the changes a 2019 German educational reform brought about for students, for instance, higher values for basic requirements or higher allowances. On the contrary, false information conditions declared that the German federal government is about to rewind the 2019 reform in the next months and as a consequence reduce monetary payments to students. In order to ensure that this information is easy to distinguish as being false or correct, a pre-test was conducted at the end of May 2020 ($N = 47$). False and correct text versions for five different topics (Port of Hamburg, cigarette advertising, housing market policy, 2% NATO defense budget goal, and the 2019 German education reform) were prepared and randomly assigned to be evaluated by participants in terms of attributed message credibility. These topics were chosen because they are known to the German public but are not so newsworthy that there would have been a risk of daily changes in the information available about them what possibly could have jeopardized the further investigation process. Results of the pretest showed that the perceived difference between message credibility of false and true texts (1 – extremely incredible to 101 extremely credible) was the highest with regard to the education reform topic (mean difference of 44.14 points). Furthermore, participants' rating of the education reform's text comprehensibility proved to be good. Thus, the false and correct education policy texts were chosen to be used for the manipulation of the information's correctness. More precisely, every participant of the main study was exposed to one of these two texts (presented in form of speeches) during the experiment's stimulus video. The pretest's data are publicly available at OSF.io.

3.2.2 Source Type

The experiment's second independent variable was the source's type. Participants were assigned to one of three conditions, either showing a video of a human, a humanoid robot, or a non-human-like device reciting the pre-tested texts about the recent education reform. Because it was the study's main purpose to allow generalizable comparisons between these three source categories that go beyond comparisons of sin-

gle devices or persons, different versions of stimulus material were created for each source condition. This means that each source type condition included two to three different speakers, and participants were randomly assigned to watch a video of only one of these speakers. For instance, participants in the robot conditions saw one of the two social, humanoid robots, Pepper or Nao (see Fig. 1). Both have a similar appearance (e.g., white enclosure, anthropomorphic embodiment), are frequently used in HRI-studies, and receive similar mediocre human-likeness scores [22]. Despite these similarities and the fact that both of them are representative examples of humanoid robots, they are still different products. The goal was not to compare these different robot exemplars but to treat both of these versions of a humanoid robot as part of the robot condition. This was done to better address the variance of this product category and thus to allow for more generalization of our findings. The second source condition (non-human-like devices) either showed smart speaker Amazon Echo Show or Google Nest Mini talking about the education reform (see Fig. 1). As in the robot condition, participants did only see one of these two smart speakers if they were assigned to the smart speaker condition. Both do not have any physical anthropomorphic design elements and would receive a human-likeness score close to zero considering the categorization of Phillips et al. [22]. Although these two smart speakers are not shaped identically due to the first one having an integrated screen for visual information, we included both as different smart speaker versions in the same condition. Again, this was done to take into account the variety of smart speaker products, equivalently to the robot condition. During the study, the smart speaker's screen, as well as the robot's screen, were only used to show an image of an audio symbol. All devices (robots and smart speakers) were programmed to communicate either the correct or false information, while the robots additionally made use of supporting human-like hand or head gestures. Lastly, in the human source condition participants saw a video with one out of three possible persons as a speaker (two women and one man in their twenties), who willingly agreed to be filmed while reciting the pre-tested information in a neutral manner. A third human was recorded to take into account the higher diversity of humans compared to technological devices. It is important to mention that all robots, smart speakers, and persons used their own voice/default voice to communicate the pretested information. In sum, three different source types (human, humanoid robot, and smart speaker) were used in the experiment, each with two to three different versions (7 in total) that were bundled together into the three conditions. In combination with the information's correctness manipulation (correct vs. false information), this resulted in six experimental conditions.

3.3 Measurements

3.3.1 Source Credibility

Source credibility was assessed via the Source Credibility Measures by McCroskey and Teven [13], measuring the three credibility subconstructs competence ($\alpha = .84$), caring/goodwill ($\alpha = .73$), and trustworthiness ($\alpha = .80$) with six items each on a seven-point semantic differential (e.g., “untrustworthy”–“trustworthy”).

3.3.2 Attribution of ToM Abilities

ToM attributions were measured by an adapted scale proposed by Benninghoff, Kulms, Hoffmann, and Krämer [32]. The two subscales awareness ($\alpha = .72$ to $\alpha = .78$) and unpredictable, with eleven items in sum (e.g., “The human/robot/smart speaker is aware of the fact that people have own wishes, thoughts and feelings.”) were measured by a seven-point Likert-scale (1—don't agree at all to 7—totally agree). However, the unpredictable subscale did not receive acceptable reliability results even after considering excluding specific items ($\alpha = .10$ to $\alpha = .41$). Because of this and the higher fit of the awareness subscale than the unpredictable subscale to our experiment (a short-term non-interactive setting) only the awareness subscale was used for further analyses including attributed ToM abilities.

3.3.3 Further Influencing Factors

Reliance on technology was measured on a seven-point Likert-scale (1—don't agree at all to 7—totally agree) using twelve items (such as “Using Information technology makes it easier to do my work.”; $\alpha = .84$; [44]). The same seven-point Likert-scale (1—don't agree at all to 7—totally agree) was used again to measure the machine heuristic's strength (four items e.g., “If a machine does a job, then the work was error-free.”; $\alpha = .73$; [39]). Moral agency attributions [34] including the two subscales morality (six items e.g., “This human/robot/smart speaker has a sense for what is right and wrong.”; $\alpha = .93$) and dependence (four items e.g., “This robot/smart speaker's actions are the result of its programming.”; $\alpha = .77$) were measured via seven-point Likert-scale (1—don't agree at all to 7—totally agree). Furthermore, the anthropomorphism subscale of the Godspeed Questionnaire was embedded in the robot and smart speaker conditions to check for differences in human-likeness in line with the manipulation using a semantic differential (five items, e.g., 1—“machinelike” to 5—“humanlike”; $\alpha = .84$; [45]).

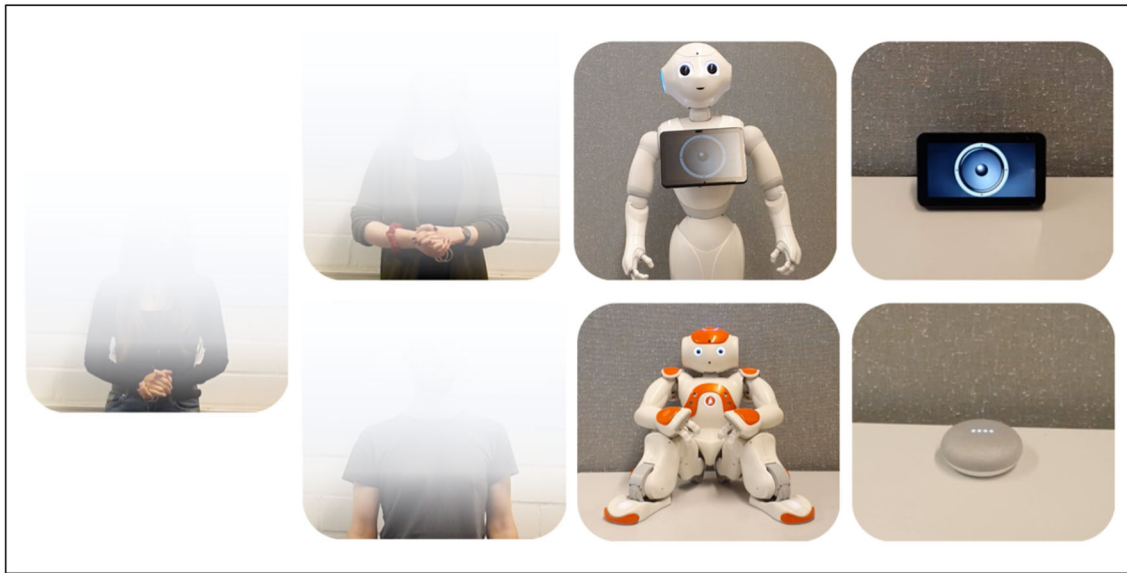


Fig. 1 Cut to size frames from the stimulus videos showing all speakers included for the manipulation of the independent variable source type. From left to right: three humans (two female, one male), two social humanoid robots (Pepper (top) and Nao (bottom)), two smart speak-

ers (Amazon's Echo Show (top right) and Google's Nest Mini (bottom right)). For publication, a blur effect was added to make the identities of the actors unrecognizable. Participants saw the persons without the blur effect

3.3.4 Additional Measures

Participants' individual issue importance was measured via a semantic differential (four items e.g., 1—"irrelevant" to 7—"relevant"; $\alpha = .91$; [39,46]) and their self-estimated issue knowledge was measured by a single-item question ("How extensive is your knowledge on the topic of state education spending?"; 1—very low pronounced to 7—very high pronounced). Finally, the degree to which participants rated the presented information credible or not (1—extremely incredible to 101—extremely credible) in combination with questions about socio-demographic information and familiarity with the sources were queried. In addition, control questions and manipulation checks were integrated, to ensure a higher quality data set.

3.4 Procedure

Participants were recruited through various free channels like social media (e.g., Facebook, Instagram, Xing, Nebenan.de), recruitment pages (e.g., Surveycircle, Poll-pool), and printed flyers. The study was accessible via a computer or mobile device. Requirements for taking part in the study included being at least 18 years old, having a good understanding of the German language as well as being in a quiet environment during the course of the study. After participants were randomly assigned to one of the six conditions, they were informed about the importance of voluntariness and anonymity during data processing. A short video with audio

instructions followed to let the participants adjust their audio settings for the upcoming video stimulus. Participants were then told to listen carefully to the respective stimulus video and to remember the communicated information to be able to answer questions about the video's content afterward. Then the video (~60s) either featuring a human, robot, or smart speaker communicating either false or correct information had to be started and watched. On the next page, three questions about the information presented during the video were asked. The following pages of the questionnaire included all self-report scales including the measurements for source credibility and attributed ToM abilities. On the last page, participants were debriefed about the experiment's purpose. In particular, participants in conditions with incorrect information were explicitly informed about the speaker's false information and were given the correct information instead. Finally, as a reward for taking part in the study, participants could sign up to take part in a voucher raffle (total value of 200€) by separately entering their email addresses. Winners were chosen by chance and contacted after the study was completed. On average, participants needed about nine minutes to complete the study ($M = 9.14$, $SD = 1.94$).

3.5 Sample

All in all, 425 people completed our online experiment. Based on analyses of required time and comparisons to normal distributions, 22 participants were excluded from analyses because they needed less than 4.17 minutes to com-

plete (lower boundary of a 95% interval centered around the raw data's mean of $M = 8.72$ ($SD = 2.31$). In addition, these 22 participants were also excluded due to failed manipulation or control checks together with 62 other persons (see Sect. 4.1). Finally, some participants in the respective conditions stated to know the persons (2,5%), the robots (4%), or the smart speakers (14%) used in the experiment. The highest share is accounted for by the smart speaker Google Nest Mini (22%). Although it is comprehensible for participants to know the robots or smart speakers since they are mass-produced goods, knowing the actors in the videos might caused unintended effects of prior knowledge. Therefore, we decided to exclude these 2.5% (three persons in absolute numbers). Thus, the study's final sample consisted of $N = 338$ participants, distributed approximately equally among the six conditions (ranging from $n = 53$ to $n = 61$). 68% of these participants were female, 31% male, and 1% divers. The participants' mean age was $M = 28.01$ ($SD = 8.58$; $M_{Median} = 25.00$, $M_{Modus} = 24$) ranging between 18 and 73 years. The majority stated to be students (65%). Furthermore, the percentage of people holding an academic degree was high (62%). The topic's importance was rated high by the participants ($M = 5.81$ $M_{Median} = 6.00$, $M_{Modus} = 7$) while self-estimated issue knowledge was medium-high/low and more normally distributed around the mean ($M = 4.12$ $M_{Median} = 4.00$, $M_{Modus} = 5$).

4 Results

4.1 Manipulation and Control Checks

Four measures were taken into account to exclude inattentive participants from the final data sample. Two manipulation checks were integrated that asked participants to choose what type of speaker delivered the information to them (woman, man, robot, smart speaker, smartphone, tv, don't know) and what has been the speaker's main topic during the video (reversal, lack, or benefits of a recent education reform). In addition, a control item ("If you are paying attention to the study, please select 'disagree' here") and a control question ("Have you filled out the survey conscientiously [...]?") supplemented the manipulation checks. After checking these criteria, 62 participants were excluded from further analyses, resulting in the final sample size of $N = 338$. All raw, as well as prepared data files, are publicly available at OSF.io.

4.2 Credibility Evaluations Based on Information's Correctness and Source Type

The first two hypotheses, H1 and H2, assumed that the level of credibility assigned to a speaker would differ depending on the type of source and on the correctness of its given informa-



Fig. 2 Descriptive mean values of all credibility subconstructs (competence, trustworthiness, goodwill) for correct information conditions (* = Significant differences to the human condition based on a 95% level of significance.)

tion. More precisely, it was assumed that machinelike sources would receive higher attributions of credibility in correct-information conditions and that human(-like) speakers would receive better ratings in false-information conditions. To test these hypotheses a multivariate analysis of variance was conducted including the two experimental variables (source type and information's correctness) as independent factors and all three measured dimensions of source credibility (competence, goodwill, and trustworthiness) as dependent variables. In Figs. 2 and 3 the credibility subscales' mean values for each condition are presented. Results revealed a significant main effect of the source type on people's credibility evaluations (human > robot, smart speaker) ($F(6, 660) = 24.25$, $p < .001$; Wilk's $\lambda = 0.671$, partial $\eta^2 = .18$).¹ Subsequent post-hoc Tukey-HSD comparisons identified that differences between source types could be explained by higher ratings for humans regarding trustworthiness ($p < .001$) and goodwill ($p < .001$) compared to the smart speakers and robots each. With regard to attributed competence humans received slightly better ratings than robots ($p = .032$, mean difference = 0.36, 95%-CI [0.02-0.70]) while the difference between humans and smart speakers was not significant ($p = .227$). However, neither the main effect of the information's correctness ($F(3, 333) = 1.97$, $p = .118$; Wilk's $\lambda = 0.982$,

¹ The study's aim was to compare social, humanoid robots against humans and non-human-like devices. Thus, we did not focus on differences between our two humanoid robots. However, we uploaded the results of additional statistical comparisons testing for differences between them on OSF.io



Fig. 3 Descriptive mean values of all credibility subconstructs (competence, trustworthiness, goodwill) for false information conditions (* = Significant differences to the human condition based on a 95% level of significance.)

partial $\eta^2 = .02$) nor the interaction of both factors (source type * correctness of information: $F(6, 666) = .954, p = .456$; Wilk's $\lambda = 0.983$, partial $\eta^2 = .01$) became significant during the analyses. Therefore, the results do not support the assumed information's correctness' influencing effect.

Additional analyses of variance revealed that the degree to which participants rated the devices' human-likeness/machine-likeness did not exactly correspond to the experiment's manipulation. All four devices (robots and smart speakers) were rated rather machine-like ($M_{Nao} = 1.40$, $M_{Pepper} = 1.55$, $M_{Google} = 1.39$, $M_{Echo} = 1.98$) with only the Echo smart speaker showing slightly higher values ($F(3, 224) = 13.31, p < .001$; partial $\eta^2 = .15$). Thus, hypotheses H1 and H2 cannot be supported based on the empirical findings. Credibility ratings (measured via its subconstructs competence, trustworthiness, and goodwill) were almost always higher for the human speakers compared to robots and smart speakers and also independent of the communicated information's correctness.

4.3 Attribution of ToM Abilities

Hypothesis H3 postulated that people attribute differently high expressed theory of mind-abilities to humans, humanoid robots, and non-human-like devices. More specifically, it was expected that the more human(-like) the subject/object in question is, the higher the level of attributed ToM abilities. To test this assumption a one-way ANOVA was calculated using the independent variable source type and the attributed ToM abilities (awareness subscale) as the dependent variable.

A significant main effect of source type was found ($F(2, 335) = 85.76, p < .001$, partial $\eta^2 = .34$). While subsequent post-hoc analyses revealed no significant differences ($p = .498$) between humanoid robots ($M = 2.26, SD = 0.97$) and smart speakers ($M = 2.42, SD = 1.09$) the human speakers ($M = 3.91, SD = 1.05$) received significantly higher ratings than robots ($p < .001$, mean difference = 1.65, 95%-CI [1.32-1.98]) and smart speakers ($p < .001$, mean difference = 1.50, 95%-CI [1.17-1.82]). Therefore, H3 receives partial empirical support as people attributed significantly higher ToM abilities to humans compared to lower and similar ratings for both, humanoid robots and non-human-like smart speakers.

4.4 Potential Moderating Influence

The last deduced hypothesis (H4) assumed that a communicated information's correctness can affect the degree to which attributed ToM abilities can influence people's credibility ratings of a technological source. For this reason, moderated linear regressions were calculated. The information's correctness served as the moderator on the influence of attributed ToM abilities (predictor) on credibility-ratings including competence, trustworthiness, and goodwill as criteria variables. Small influences of attributed ToM abilities on people's ratings of devices' competence ($\beta = .216, p = .001$), trustworthiness ($\beta = .18, p = .007$), and goodwill ($\beta = .274, p < .001$) were detected. Equivalently to the non-significant main effect of the information's correctness on credibility attributions, during none of these analyses a moderating effect of the information's correctness became significant ($p = .528$ to $p = .676$). In order to not interpret the influence of attributed ToM abilities without context, we decided to integrate them as a predictor into the multiple regression analyses which were calculated for Research Question 2. With regard to H4, however, no empirical support based on the collected data can be provided.

4.5 Exploratory Research Questions

The first research question tried to address the influence of the machine heuristic on the evaluation of the information's credibility itself. Therefore, correlations between these two variables were calculated for conditions including technological devices (robot and smart speakers). No significant correlations between the machine heuristic's strength and the extent to which participants judged the communicated information to be credible could be discovered ($r = .04, p = .563$). Furthermore, no differences in strength of the heuristic between groups with false and true information communicated by robots or smart speakers were found ($p = .628$). Thus, it seems unlikely to assume an influence of people's thoughts about the information impartiality of machines on

the perception of the experiment's manipulated information's credibility.

In order to properly address the second research question, multiple regression analyses were performed. The aim was to gain a better understanding of different constructs' influences on robot-related credibility attributions, such as morality, dependence, reliance on technology, strength of the machine heuristic, and attributed degree of anthropomorphism. Multiple regression analyses revealed people's attributed anthropomorphism to be one of the most influential predictors for each of the robots' credibility subconstructs, including competence ($\beta = .39$), trustworthiness ($\beta = .33$), and goodwill ($\beta = .24$). Also, other constructs such as people's reliance on technology ($\beta_{competence} = .26$, $\beta_{trustworthiness} = .30$, $\beta_{goodwill} = .16$) and their attributed morality ($\beta_{competence} = .13$, $\beta_{trustworthiness} = .28$, $\beta_{goodwill} = .27$) showed to have influence on credibility attributions. With focus on the subconstruct goodwill, also the machine heuristic showed some small negative influence ($\beta_{competence} = .14$, $\beta_{trustworthiness} = -.01$, $\beta_{goodwill} = -.20$). Attributed ToM abilities on the other hand did not turn out to be a prominent predictor of the robots' credibility ratings ($\beta_{competence} = .06$, $\beta_{trustworthiness} = -.02$, $\beta_{goodwill} = .14$) when being included in these multiple regression analyses.

5 Discussion

All in all, the purpose of this research work was to investigate the influence of different source types (human, humanoid robot, non-human-like devices) on people's respective credibility attributions depending on either correct or false communicated information. As outlined in the beginning, the usage of social robots in information-sharing tasks may increase continuously in the near future and will thus raise the importance to understand if and how robotic museum guides, assistants in elderly care, or robots as teachers are judged as credible sources of information. Based on theoretical considerations related to the embodiment's influence in human-robot interactions, it was assumed that the human-likeness of robots would be able to negate adverse effects of algorithm aversion which can occur when algorithmically produced information is false [7]. However, this was not the case in our experiment. Regardless of the information's correctness, human sources always received significantly higher credibility ratings than the robots, as well as the smart speakers did, while robots and smart speakers barely differed. The hybridity of the robots' anthropomorphic physiques did neither let them be evaluated similar to their human counterparts nor did it let them be evaluated better than non-human-like devices after communicating false information. Our findings

receive additional support by the first research question's results which did not detect any relationship between the machine heuristic's strength and the degree to which participants perceived the presented information to be correct. Nevertheless, this analysis has to be treated with care because the information's correctness was varied as the independent variable in this experiment and was only measured as an additional control question. Furthermore, partly in line with the derived hypothesis, the humanoid robots as well as the smart speakers received lower attributions of ToM abilities. Although these attributions showed to be slightly influential with regard to the sources' credibility evaluations, no moderating impact of the information's correctness was found. Furthermore, the final exploratory, multiple regression analyses additionally revealed that especially people's reliance on technology, their rating of a robot's degree of anthropomorphism, and the attributed level of morality were the most influential predictors with regard to the robots' credibility ratings.

In sum, the calculated results implicate that people are not simply granting a humanoid robot more credibility because of its anthropomorphic exterior. The significant main effect of source type supports the assumption that humans are in favor of being perceived in part more competent but especially with regard to soft skills more trustworthy and good willing than technological devices independent from the communicated information's correctness. Although differences in competence ratings between smart speakers and human sources failed to reach this significance the descriptive means also fit the trend of generally higher credibility attributions for humans. These findings are in line with the conclusions of Waddell [39], who discovered that news articles are perceived less credible when a machine instead of a human is listed as its author. Nonetheless, a similar interpretation of results seems difficult because Waddell [39] addresses the problem of expectancy violations due to a task (journalism) mostly associated with humans. This explanation does not seem to fit here because for example smart speakers are already widely used for delivering news information but were rated less credible than humans regardless of the condition. In the following, we want to discuss two other possible explanations for these findings. The first one takes into account the probability that the social cues offered by the robots' anthropomorphic embodiments (e.g., facial features such as eyes or body features such as arms and a human-like torso) need to exceed a certain threshold to trigger credibility attributions that could be observed for the human sources. This relates to the principle of similarity [47] which describes that computer products that are more similar to their users will be rated as more credible and thus be more persuasive in interactions. Although both robots used in the study (Nao and Pepper) are common

versions of human-like robots to encounter in research and public, they only represent an intermediate level of physical human likeness [22] and therefore might not have been able to exceed this threshold. Our analyses support this idea, as they revealed that although participants' ratings of anthropomorphism were an influential predictor of credibility attributions related to robots, both robots were rated rather machine-like than human-like. From this point of view, it seems that more realistically human-like robotic devices are needed (e.g., a more similar surface look including skin) in order to observe more human-like credibility attributions. However, given that more realism in surface features does not necessarily correlate with better evaluations of robots [38] and that people are often willing to surrender to a willing suspension of disbelief eliminating the need for a completely perfect realistic design [48], a more detailed anthropomorphically embodiment does not necessarily have to be a promising approach. On the contrary, a second explanation for our results includes the assumption of a human superiority effect. It could be argued that technological devices no matter how anthropomorphically they are embodied will always receive lesser credibility attributions than humans will. Although competence ratings did not differ heavily between technological devices and humans, clear distinctions with regard to the credibility sub-constructs trustworthiness and goodwill could be detected. A possible explanation for these results is the fact that competence attributions are more based on the physical ability to possess meaningful knowledge, while trustworthiness and goodwill are rather social qualities. These social attributions might not be applicable to technological devices to the same degree as they are to humans and therefore cause humans to be attributed more credible in sum. This conclusion receives support from the different degrees of ToM abilities attributed to humans, robots, and smart speakers. However, this result in part contradicts previous research results which found a difference in mental state attributions depending on the degree of human-likeness of technological devices [36,37]. Again, humans received higher ratings than both, smart speakers and robots, in our experiment. With regard to credibility attributions, it seems reasonable to argue that in a first step a certain degree of consciousness, as well as the ability to understand the existence of feelings and needs in other persons, are mandatory prerequisites for a technological device in order to be even able to be attributed good willing in a second step. Fitting to this assumption attributed ToM abilities especially influenced ratings of the robots' goodwill. However, our experiment did not include any cues to actively support the attribution of ToM abilities, but only measured people's general likeliness of attributing them to a device they only experienced through video interaction. Thus, the overall low attributions of ToM abilities might have been one further reason why robots were not rated more credible, especially good willing. The positive effect of perceived

morality further supports this argument and demonstrates the need to think of attributions of mental capacities as a relevant influencing factor for robotic credibility attributions.

Another important finding was the non-significant main effect of the information's correctness and its lack of interaction with the independent variable source type. Based on these results it should not mistakenly be argued that correct and false information always causes similar source credibility attributions. Nevertheless, no differences in attributed source credibility could be detected in our study despite the supporting pre-test results and the topic's noticeable high importance for the participants. One explanation for this non-significance could be that participants did not focus on the communicated information as much as they did on the source itself (human, robot, smart speaker). This line of reasoning does not seem satisfactory because all analyses strictly included only participants which paid attention to the stimulus information. Therefore, a probably more fitting argument could be that the communicated information was too impersonal to be relevant despite the topic's general importance. During the study no participant had to fear immediate consequences from relying on any false information e.g., receiving a lower financial compensation. Thus, the information's correctness was not decision-relevant at any point during the study which could have caused participants to not consider its value for their credibility attributions. To investigate this assumption, future studies should analyze the influence of an information's decision-making relevance on the relation between the information's correctness and robotic source credibility attributions.

In addition to all these findings, the influence of user-related constructs cannot be neglected in the credibility attribution process, as reliance on technology has been shown to be one of the most promising predictors of participants' robot credibility evaluations. Previous studies have already shown that technology-related user characteristics like technical affinity can be meaningful predictors in human-robot interactions [41] which is supported by our findings. It seems plausible and consistent that higher levels of reliance on technological products correlate with higher credibility ratings maybe because of generally higher positive experiences with these devices.

Based on the experiment's design, all discussed insights are especially important to consider during short-term human-robot interactions with unidirectional information-sharing tasks. Because the effects were analyzed not only for one specific robot or smart speaker product, it seems reasonable to expect that similar effects and restrictions concern comparable social, moderately humanlike robots in equal manners. However, it is important to keep in mind that this group of moderately human-like robots should not be seen as completely identical with regard to every aspect of people's perceptions of them because even in this subgroup of robots

internal differences can be found [25]. Concepts for successful human-robot interactions should therefore consider if a specific humanoid robot is suited for the respective task and will not be negatively affected by (at least) initially lower credibility attributions. Thus, tasks that are purely about communicating information, for instance in a shop or a video, might be less suitable usage contexts for humanoid robot's, especially if the interaction time is limited and human-like abilities such as touch, gestures, or gaze are not necessarily needed or cannot be used. In addition, such low engaging interactions might not be sufficient to support people's attribution of ToM-abilities towards robots compared to more interactive encounters which could reveal such attributions [37]. These problems might be countered by individuals who show higher reliance on technological products in general, but because relying on robots in particular is not commonplace, the discussed problems should be taken into account when creating successful human-robot interactions independent of potentially helpful user characteristics.

5.1 Limitations

When discussing limitations to this study it has to be noted that due to the study's convenience sample, it included above-average female and higher-educated participants. Furthermore, the experimental manipulation only made use of video stimulus material, that was created in ways to be similar between all six conditions. Everyday human or human-robot interactions, however, are far more complex and include components that are not applicable and therefore difficult to compare to communications with smart speakers, for instance, the usage of locomotive features during interactions or the possibility to make use of touch. It might be true that more profound interactions are needed which are not only unidirectional and based on a short information sharing scene. Thus, future investigations on longer-term and more dyadic human-robot interactions would be helpful to gain more ecologically valid results. In addition, it might also be difficult to assess credibility attributions via the same measurement instrument for humans and technological devices. Although the Source Credibility Measures [13] have already been used to measure an artificial entity's credibility in many experiments (see for example [5,28]), it was not constructed and validated for this research purpose. Items like "has my interests at heart" might be more applicable to human sources because they were developed and phrased for human interaction partners in the first place. In combination with the difficulty of reliable, directly measured self-reports, future studies should ideally try to replicate our results with even more comparable and indirect measurements of credibility attributions such as usage rates of decision-relevant communicated information.

5.2 Conclusion

All in all, the experiment's aim was to investigate differences and similarities in credibility attributions related to humans, humanoid robots, and non-human-like devices. Our results show that humans turned out to be the most credible sources, while all technological devices (robots and smart speakers) were rated especially less trustworthy and caring/good willing in comparison. Furthermore, higher attributions of theory of mind abilities to human sources without significant differences between humanoid robots and smart speakers were discovered. Although these ToM attributions slightly influenced the robots' credibility evaluations, more prominent influential factors turned out to be attributed anthropomorphism, moral agency, and people's general reliance on technology. In sum, this study concludes that mediocreatly anthropomorphically embodied robots, but also non-human-like devices might not be credible sources of information by default if being compared to human sources in short-term, unidirectional, and mere information delivering interactions.

Funding Open Access funding enabled and organized by Projekt DEAL. The study only received internal financial support from the budget of the social psychology chair (University of Duisburg-Essen) where the experiment was conducted.

Data Availability The study project, its preregistration and data files are publicly available at: <https://osf.io/hvnft/>.

Declarations

Conflict of interest There are no conflicts of interest or competing interests with regard to the study.

Ethical Approval An approval of the university's ethical committee was obtained before the study started (ID: 2008SPFM3311).

Consent to Participate Participants were informed about the study's procedure on the first page of the online questionnaire and that they should only begin the experiment if they agree to the conditions described there.

Consent to Publication All three persons seen in the experiment's stimulus material consented to the use of this material for the study and to associated publications.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Baraka K, Alves-Oliveira P, Ribeiro T (2019) An extended framework for characterizing social robots. *arXiv*. <https://arxiv.org/abs/1907.09873>. Accessed 29 July 2021
- Pornpitakpan C (2004) The persuasiveness of source credibility: a critical review of five decades' evidence. *J Appl Soc Psychol* 34(2):243–281. <https://doi.org/10.1111/j.1559-1816.2004.tb02547.x>
- Metzger MJ, Flanagin AJ (2013) Credibility and trust of information in online environments: the use of cognitive heuristics. *J Pragmat* 59(B):210–220. <https://doi.org/10.1016/j.pragma.2013.07.012>
- Cho E, Sundar SS, Abdullah S, Motalebi N (2020) Will deleting history make alexa more trustworthy? Effects of privacy and content customization on user experience of smart speakers. In: Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems (CHI'20) pp 1–13. <https://doi.org/10.1145/3313831.3376551>
- Edwards A, Edwards C, Spence PR, Harris C, Gambino A (2016) Robots in the classroom: differences in students' perceptions of credibility and learning between "teacher as robot" and "robot as teacher". *Comput Hum Behav* 65:627–634. <https://doi.org/10.1016/j.chb.2016.06.005>
- Sundar SS (2008) The MAIN model: a heuristic approach to understanding technology effects on credibility. In: Metzger MJ, Flanagin AJ (eds) *Digital media, youth, and credibility*. The MIT Press, Cambridge, pp 73–100
- Dietvorst BJ, Simmons JP, Massey C (2015) Algorithm aversion: people erroneously avoid algorithms after seeing them err. *J Exp Psychol Gen* 144(1):114–126. <https://doi.org/10.1037/xge0000033>
- Fink J (2012) Anthropomorphism and human likeness in the design of robots and human-robot interaction. In: Ge SS, Khatib O, Cabibihan JJ, Simmons R, Williams MA (eds) *Social Robotics. ICSR 2012. Lecture Notes in Computer Science*, vol 7621. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-34103-8_20
- Seiler R, Kucza G (2017) Source credibility model, source attractiveness model and match-up-hypothesis: An integrated model. *J Int Sci Publ Econ Bus*. <https://doi.org/10.21256/zhaw-4720>
- Hwang S (2013) The effect of Twitter use on politicians' credibility and attitudes toward politicians. *J Publ Relat Res* 25(3):246–258. <https://doi.org/10.1080/1062726X.2013.788445>
- Brodsky SL, Griffin MP, Cramer RJ (2010) The witness credibility scale: an outcome measure for expert witness research. *Behav Sci Law* 28(6):892–907. <https://doi.org/10.1002/bsl.917>
- Kohring M (2001) Vertrauen in Medien—Vertrauen in Technologie. <https://elib.uni-stuttgart.de/handle/11682/8694>. Accessed 29 July 2021. <https://doi.org/10.18419/opus-8677>
- McCroskey JC, Teven JJ (1999) Goodwill: a reexamination of the construct and its measurement. *Commun Monogr* 66(1):90–103. <https://doi.org/10.1080/03637759909376464>
- Fogg BJ, Tseng H (1999) Social facilitation with social robots? In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, 80–87. <https://doi.org/10.1145/302979.303001>
- Reeves B, Nass CI (1996) *The media equation: how people treat computers, television, and new media like real people and places*. Cambridge University Press, Cambridge
- Zlotowski J, Sumioka H, Eyssel F, Nishio S, Bartneck C, Ishiguro H (2018) Model of dual anthropomorphism: the relationship between the media equation effect and implicit anthropomorphism. *Int J Soc Robot* 10:701–714. <https://doi.org/10.1007/s12369-018-0476-5>
- Nass C, Moon Y (2000) Machines and mindlessness: social responses to computers. *J Soc Issues* 56(1):81–103. <https://doi.org/10.1111/0022-4537.00153>
- Hoffmann L, Krämer NC, Lam-chi A, Kopp S (2009) Media equation revisited Do users show polite reactions towards an embodied agent. In: Ruttkey Z, Kipp M, Nijholt A, Vilhjálmsson HH (eds) *Intelligent Virtual Agents IVA 2009 Lecture Notes in Computer Science*. Springer, Berlin, Heidelberg
- Riether N, Hegel F, Wrede B, Horstmann G (2012) Social facilitation with social robots? In: Proceedings of the 7th ACM/IEEE International Conference on Human-Robot Interaction (HRI), pp 41–47. <https://doi.org/10.1145/2157689.2157697>
- Strathmann C, Szczuka J, Krämer NC (2020) She talks to me as if she were alive: Assessing the social reactions and perceptions of children toward voice assistants and their appraisal of the appropriateness of these reactions. In Proceedings of the 20th ACM International Conference on Intelligent Virtual Agents (IVA'20), 52. <https://doi.org/10.1145/3383652.3423906>
- Lemaignan S, Fink J, Dillenbourg P, Braboszcz C (2014) The cognitive correlates of anthropomorphism. EPFL Infoscience. <https://infoscience.epfl.ch/record/196441>. Accessed 29 July 2021
- Phillips E, Zhao X, Ullman D, Malle BF (2018) What is human-like? Decomposing robots' human-like appearance using the anthropomorphic robot (abot) database. In: Proceedings of the 2018 ACM/IEEE International Conference on Human-Robot Interaction, pp 105–113
- Caudwell C, Lacey C (2020) What do home robots want? The ambivalent power of cuteness in robotic relationships. *Converg* 26(4):956–968. <https://doi.org/10.1177/1354856519837792>
- Epley N, Waytz A, Cacioppo JT (2007) On seeing human: a three-factor theory of anthropomorphism. *Psychol Rev* 114(4):864–886. <https://doi.org/10.1037/0033-295X.114.4.864>
- Manzi F, Massaro D, Di Lernia D, Maggioni MA, Riva G, Marchetti A (2021) Robots are not all the same: young adults' expectations, attitudes, and mental attribution to two humanoid social robots. *Cyberpsychol Behav Soc Netw* 24(5):307–314. <https://doi.org/10.1089/cyber.2020.0162>
- Bartneck C, Forlizzi J (2004) A design-centred framework for social human-robot interaction In: Proceedings of the 13th IEEE International Workshop on Robot and Human Interactive Communication (RO-MAN 2004) pp 591–594. <https://doi.org/10.1109/ROMAN.2004.1374827>
- Andrist S, Spann E, Mutlu B (2013) Rhetorical robots: making robots more effective speakers using linguistic cues of expertise. In: Proceedings of the 8th ACM/IEEE International Conference on Human-Robot Interaction (HRI), pp 341–348. <https://doi.org/10.1109/HRI.2013.6483608>
- Spence PR, Edwards C, Edwards A, Lin X (2019) Testing the machine heuristic: Robots and suspicion in news broadcasts. In: Proceedings of the 14th ACM/IEEE International Conference on Human-Robot Interaction (HRI), pp 568–569. <https://doi.org/10.1109/HRI.2019.8673108>
- Chidambaram V, Chiang Y-H, Mutlu B (2012) Designing persuasive robots: How robots might persuade people using vocal and nonverbal cues. In: Proceedings of the 7th ACM/IEEE International Conference on Human-Robot Interaction (HRI), pp 293–300. <https://doi.org/10.1145/2157689.2157798>
- Logg JM, Minson JA, Moore DA (2019) Algorithm appreciation: people prefer algorithmic to human judgment. *Organ Behav and Hum Decis Process* 151:90–103. <https://doi.org/10.1016/j.obhdp.2018.12.005>
- Prahl A, Van Swol L (2017) Understanding algorithm aversion: When is advice from automation discounted? *J Forecast* 36(6):691–702. <https://doi.org/10.1002/for.2464>

32. Benninghoff B, Kulms P, Hoffmann L, Krämer NC (2013) Theory of mind in human-robot-communication: Appreciated or not? *Kognitive Systeme* 1. <https://doi.org/10.17185/dupublico/31357>
33. Sturgeon S, Palmer A, Blankenburg J, Feil-Seifer D (2019) Perception of social intelligence in robots performing false-belief tasks. In: Proceedings of the 28th IEEE International Conference on Robot and Human Interactive Communication, pp 1-7. <https://doi.org/10.1109/RO-MAN46459.2019.8956467>
34. Banks J (2019) A perceived moral agency scale: development and validation of a metric for humans and social machines. *Comput Hum Behav* 90:363–371. <https://doi.org/10.1016/j.chb.2018.08.028>
35. Goldman AI (2012) Theory of mind. In: Margolis E, Samuels R, Stich SP (eds) *The Oxford handbook of philosophy of cognitive science*. Oxford University Press, Oxford, New York, pp 402–424
36. 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):e2597. <https://doi.org/10.1371/journal.pone.0002597>
37. Manzi F, Peretti G, Di Dio C, Cangelosi A, Itakura S, Kanda T, Ishiguro H, Massaro D, Marchetti A (2020) A robot is not worth another: exploring children's mental state attribution to different humanoid robots. *Front Psychol* 11:2011. <https://doi.org/10.3389/fpsyg.2020.02011>
38. Rosenthal-von der Pütten AM, Schulte FP, Eimler SC, Sobieraj S, Hoffmann L, Maderwald S, Brand M, Krämer NC (2014) Investigations on empathy towards humans and robots using fMRI. *Comput Hum Behav* 33:201–212. <https://doi.org/10.1016/j.chb.2014.01.004>
39. Waddell TF (2018) A robot wrote this? How perceived machine authorship affects news credibility. *Digit J* 6(2):236–255. <https://doi.org/10.1080/21670811.2017.1384319>
40. Sundar SS, Kim J (2019) Designing persuasive robots: How robots might persuade people using vocal and nonverbal cues. In: Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems, 538. <https://doi.org/10.1145/3290605.3300768>
41. Horstmann AC, Krämer NC (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. *Front Psychol* 10:939. <https://doi.org/10.3389/fpsyg.2019.00939>
42. Deley T, Dubois E (2020) Assessing trust versus reliance for technology platforms by systematic literature review. *Soc Med Soc* 6(2):1–8
43. Jackson RB, Williams T (2019) On perceived social and moral agency in natural language capable robots. In: Proceedings of the 2019 HRI Workshop on the Dark Side of Human-Robot Interaction, pp. 401-410
44. Marathe SS, Sundar SS, Bijvank MN, van Vugt H, Veldhuis J (2007) Who are these power users anyway? Building a psychological profile. Paper presented at the 57th Annual Conference of the International Communication Association, San Francisco
45. Bartneck C, Kulić D, Croft E, Zoghbi S (2009) Measurement instruments for the anthropomorphism, animacy, likeability, perceived Intelligence, and perceived safety of robots. *Int J Soc Robotics* 1:71–81. <https://doi.org/10.1007/s12369-008-0001-3>
46. Paek HJ, Hove T, Kim M, Jeong HJ, Dillard JP (2012) When distant others matter more: perceived effectiveness for self and other in the child abuse PSA context. *Media Psychol* 15(2):148–174. <https://doi.org/10.1080/15213269.2011.653002>
47. Fogg BJ (2002) *Persuasive technology: Using computers to change what we think and do*. Morgan Kaufmann, San Francisco
48. Duffy BR, Zawieska K (2012) Suspension of disbelief in social robotics. In: Proceedings of the 21st IEEE International Symposium on Robot and Human Interactive Communication (2012 IEEE RO-MAN), pp 484-489. <https://doi.org/10.1109/ROMAN.2012.6343798>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Marcel Finkel is working at the University of Duisburg-Essen (Germany) as a research associate at the department of “Social Psychology: Media and Communication” since 2020, where he is working on his Ph.D. focused on human-robot-interaction research.

Nicole C. Krämer is Professor for “Social Psychology: Media and Communication” at the University of Duisburg-Essen (Germany) since 2007. Her research interests include human-computer-interaction (including social robots and virtual agents) and computer-mediated communication. <https://www.uni-due.de/sozialpsychologie/kraemer.shtml>.