Abstract— A recent theoretical framework on anthropomorphism emphasizes the role of elicted agent knowledge in anthropomorphic inferences about nonhuman entities. According to the Three-Factor Model of psychological Anthropomorphism, people use anthropocentric knowledge structures when judging unfamiliar objects (e.g., robots). In the present research, our goal was to manipulate the accessibility of such elicted agent knowledge by varying features of a robot’s voice: Specifically, we examined effects of vocal cues that reflected both gender of robot (i.e., a male vs. female voice) and voice type (i.e., a human-like vs. robot-like voice). This was done to test the impact of these vocal features on anthropomorphic inferences about the robot and on human-robot interaction (HRI) acceptance. Our results demonstrate that a robot’s vocal cues clearly influence subsequent judgments of the robot and particularly so, when participant gender taken into account. Implications of our research for robotics will be discussed.

I. INTRODUCTION

When meeting a person for the first time, we nearly automatically rely on visual cues that indicate social group membership. For instance, a person’s face provides information about the individual’s age, gender or ethnicity. Ample social psychological research has demonstrated that we quickly categorize a person in terms of being a member of a social group and we thus judge the newly encountered individual in accord with the activated knowledge structures (e.g., [1-5]). This helps us to form impressions instantly.

Naturally, there are additional and even quite subtle cues on which we rely when judging others. However, whereas social psychological literature on effects of visual cues of group membership on person perception is extensive, much less is known about effects of auditory cues on impression formation and social judgments [6] in human-human context, but likewise in robotics.

Imagine setting up your assistant robot at home – in what type of voice should it respond to your requests? Should it convey the crucial information using a human-like or rather a synthesized voice? Should the machine’s voice indicate a certain gender, and if so, do vocal gender cues change user perceptions, specifically if the robot’s gender matches the user’s own?

This research aims at providing empirical answers to these research questions by embedding the current work in a recent theoretical framework, the Three-Factor Model of Anthropomorphism [7]: We assume that by means of the implementation of vocal cues (e.g., a human-like, same-gender voice) in a robot, knowledge structures related to the broader human category become activated. This way, we operationalize the notion of ‘elicted agent knowledge’ that is proposed by [7] (see Section II). The activation of ego- and anthropocentric knowledge (i.e., knowledge about oneself and one’s own human reference group) should bias social judgments: Specifically, in the context of the present research, we would predict that increasing the accessibility of elicted agent knowledge should result in stronger anthropomorphic inferences and human-robot acceptance than in the control group. Summing up, we argue that participants would project typically human traits to a system that appears similar to themselves on the basis of the system’s voice cues and this would also show in increased liking and human-robot interaction (HRI) acceptance.

II. RELATED WORK

In their classic “computers as social actors” (CASA) approach, [8-10] have already demonstrated that people interact with computers in ways that are comparable to human-human interaction and usually, they do not even give it a second thought. To illustrate, Nass and colleagues have found that people instinctively treat computers like humans, for instance, by mindlessly applying human social categories to computers, such as ethnicity or gender to the machine [8-10]. The results from HCI have been replicated in the area of social robotics (for an overview, see [11]; see also [12, 13]). In this context, too, cues conveyed by a technical system increase the accessibility of category-relevant knowledge structures which in turn serve as a basis for inferences about a robot’s personality and functions. For instance, Powers and colleagues [14, 15] have proposed that from the reliance on such cues, efficient HRI and communication processes emerge. They argue that this is due to the common ground that is established between the robot and the user (see also [13]).

More recent theorizing about psychological anthropomorphism is in keeping with the reasoning by
Powers and others [14, 15]. Specifically, Epley and colleagues have developed the Three-Factor Model of Anthropomorphism [7], which rests on the assumption that the extent to which people anthropomorphize objects and nonhuman agents can be largely attributed to three key psychological factors: sociality motivation, effectance motivation, and elicited agent knowledge. Whereas the first two factors are motivational in nature, the latter, elicited agent knowledge, represents the cognitive determinant of anthropomorphic inferences.

In the following, the three determinants will be outlined briefly: First, sociality motivation reflects the human need and desire to establish social connections with other humans. From this follows that if people are deprived of social connection, they humanize even nonhuman entities, such as robots. This serves the function of compensating for the experienced lack of social support and resulting feelings of loneliness [16].

Second, effectance motivation, serves to satisfy the human need for mastery and control over the given social environment. From our own experience, we might know that when our self-image of being a competent and efficient social agent is threatened, uncertainty commonly arises. Epley and colleagues have argued that this emergence of psychological stress and uncertainty seems particularly likely when we encounter unknown agents, for instance, a novel technical system. As a consequence, anthropomorphic inferences about the unknown entity increase, as shown in [17-19].

Most important for the present research, however, is the idea that people are likely to anthropomorphize an unfamiliar entity because knowledge structures related either to themselves or the broader category of “humans” becomes activated and accessible. These schemata, in turn, guide subsequent information processing and agent-related judgments. In other words, people use elicited agent knowledge to form a common ground with the unfamiliar entity and can do so by attributing human characteristics to it.

We argue, that elicited agent knowledge does not only encompass knowledge structures related to the superordinate human category, but also subordinate social categories, such as ethnicity or gender. To illustrate, only recently, [19] have shown that the mere manipulation of a robot’s ethnicity by changing its name and location of production influenced the extent to which participants attributed mind and other essentially and typically human traits to the robots. Moreover, participants reported feeling closer to the robot when it ostensibly belonged to their German ingroup vs. a Turkish outgroup. Similarly, [20] have conducted an experiment using the same robot platform, but this time the robot was ‘gendered’ based on visual gender cues, namely hairstyle. In this experiment, they asked participants to evaluate ‘gendered’ robots with regard to gender-stereotypical traits (e.g., warm, trusting vs. dominant, determined; see [21, 22]). Additionally, participants evaluated the robots’ suitability for pretested stereotypically female tasks (e.g., household maintenance, patient care) and stereotypically male tasks (e.g., transporting goods, repairing technical equipment). Results showed that participants readily applied gender categories and the respective stereotypes that go along with them to humanoid robots that appeared male vs. female. Specifically, the female robot prototype was perceived as warmer than the male counterpart, whereas more competence was attributed to the latter relative to the female robot prototype.

Whereas the experiment by [20] was the first to demonstrate stereotype effects of visual gender cues in robots, the existing body of research on gender effects in social robotics heavily builds on previous research by proponents of the CASA approach and has primarily focused on a machine’s synthetic voice as the main cue to trigger gender stereotyping of machines. For example, [23] have shown that participants attributed gender to computers that communicated in a low-pitched vs. high-pitched synthetic voice. Subsequently, the low vs. high frequency of the synthetic computer voice triggered gender-schematic judgments of the ‘male’ vs. ‘female’ computer. Specifically, the female-voiced computer in a dominant role was perceived as more negatively than the male-voiced dominant computer. Furthermore, evaluations provided by the ‘male’ computer were taken more seriously than when praise was given by the ‘female’ computer.

Similar findings have been obtained in the context of social robots: To provide an example, [11] have found effects of robot gender on conversation efficiency. They proposed that communication becomes more or less efficient depending on the “persona” of the robot. To illustrate, they found that female communicated more efficiently with a same-sex robot due to assumed common ground with it, whereas they appeared to discuss issues related to dating and romantic relationships more extensively when confronted with a male robot type. The authors argue that in this case, the participants assume that the male robot knows less about dating norms than a female counterpart.

Moreover, [15] have investigated the effects of physical appearance and voice frequency on the attribution of sociability and competence to a robot. They tested the assumption that a baby-faced humanoid robot would be perceived as more sociable, but less competent than a mature-faced robot and that the robot’s appearance would influence advice-taking intentions of the perceivers. Indeed, the authors found that baby-facedness predicted perceived sociability of the robot as well as participants’ intentions to take health advice from the robot. Furthermore, low voice frequency of the robot predicted perceived knowledge of the robot and participants’ willingness to take advice from it.

In sum, it is clear from findings by [12, 13, 23-25] that effects of vocal cues on perceptions of computers as well as robots have been studied previously. However, to date, the effects of both gender of robot voice and voice type have not been tested yet simultaneously.

Crucially, effects of vocal cues have not yet been investigated and interpreted in the context of Epley’s very recent theoretical model of anthropomorphism. Thus, we manipulate both gender of robot voice and voice type, thereby operationalizing elicited agent knowledge. As argued before, the activation of elicited agent knowledge
should shape user perceptions, and even more strongly so, when user gender is considered in parallel [24, 25].

To close this research gap, we conducted an experiment to investigate in more depth the role of vocal cues indicating gender and human vs. robot-likeness, respectively. We did so by studying effects on various dimensions of psychological anthropomorphism:

Specifically, we predicted that a robot that speaks with a human voice should be anthropomorphized more strongly than a robot speaking with a synthesized, machine-like voice. Furthermore, we predicted interactions between gender of robot voice and gender of participant, as well as three-way interactions: We hypothesized such effects regarding likeability, perceived psychological closeness to the robot, and contact intentions. With these dependent measures sought to tap aspects of HRI acceptance.

To test our hypotheses regarding the activation elicited agent knowledge and its effect on social perception of a robot, we conducted an experiment.

III. Method

A. Participants and Design

58 students (31 women, 27 men; \(M_{\text{Age}} = 22.98, SD = 2.81\)) were recruited on campus at Bielefeld University.

Participants were randomly assigned to one of the experimental conditions that resulted from the 2 (gender of robot voice: male vs. female) x 2 (voice type: robotic vs. human) x 2 (gender of participant: male vs. female) between-participants design.

B. Procedure

Participants were tested in a quiet laboratory setting. For the duration of the computerized experiment, they were seated in front of individual computers with headphones.

All instructions were presented on the computer screen and participants learned that the study would examine their first impressions of a newly developed robot that was to be used as a personal assistant in the future (e.g., to provide services of a ‘time manager’). To make this cover story appear more realistic, participants were presented with a video clip of the robot. In the video sequence, the robot solely uttered the neutral sentence ‘It is quarter past three’. Therefore, the video clip lasted only several seconds. Depending on experimental condition, this sentence was uttered either in a human-like or a robot-like gendered voice.

After watching the video clip, participants completed the dependent measures that were presented in different survey ‘modules’. Importantly, in between these modules, the video clip was presented again to make sure that elicited agent knowledge was activated, depending on experimental condition. Finally, participants reported demographics and were then reimbursed, debriefed, and dismissed.

C. Hardware

In the current experiment, we used the robot Flobi [17, 20, 29] to investigate the effects of gender of robot voice and voice type on evaluations of the robot.

The robot head has 18 degrees of freedom to express emotional states, such as happiness, sadness, fear, surprise, and anger. Two actuators rotate the eyebrows, three actuators move Flobi’s eyes, four actuators move the upper and lower eyelids, three actuators move the neck, and finally, six actuators were implemented to animate the robot’s lips. Furthermore, by means of four LEDs, red or white light can be projected onto the robot’s cheek surfaces.

Importantly, in the present research, participants solely relied on vocal cues to form their impressions of the robot, because robot gender was not indicated visually. Concretely, the robot was presented without the hair module to study isolated effects of vocal cues A pretest had indicated that the prototype was perceived as gender neutral. Figure 1 depicts the social robot used in the current research.

Fig. 1. The social robot Flobi.

IV. Measures

7-point Likert scales were used to collect participants’ responses to the dependent measures. For subsequent data analysis, average scores were computed to form indices of the respective dimensions, with higher values reflecting greater agreement with the assessed dimensions.

A. Manipulation Check

Participants used 7-point Likert scales to rate the extent to which they perceived the robot’s voice as human-like (1) versus robot-like (7). Analogously, participants indicated whether the voice would sound rather female (1) or male (7). This way, we assured that the voices types were adequately recognized as human- vs. robot-like and masculine vs. feminine.

B. Likeability

Two items were administered to form a reliable likeability index of the robot prototype (‘How warmhearted / likeable do you consider the robot?’ (\(\alpha = .77\)).

C. Psychological Closeness to the Robot

To assess the degree of perceived psychological closeness between participants and Flobi, participants responded to five items: ‘To what extent do you feel close / connected / similar to the robot / on the same wavelength’ and ‘Do you share many communalities with the robot?’ (\(\alpha = .77\)).
This measure was characterized by a good internal consistency ($\alpha = .85$).

D. Contact Intentions

To measure contact intentions, we used the following items: ‘How much would you like to get acquainted with the robot?’, ‘Given you had the money, how much would you like to buy the robot?’, ‘How eager are you to have the robot at your home’, and ‘How interested are you in talking to the robot eventually?’ (see also [19]). The four items formed a reliable index of contact intentions ($\alpha = .83$)

E. Anthropomorphism

To investigate the extent to which participants would anthropomorphize the robot, we used two different measures:

Firstly, to assess participants’ anthropomorphistic inferences about the robot, we presented them with 10 personality traits that reflect human essence [26, see also 17]. Specifically, the selected traits represented the dimension of ‘human nature’.

Human nature traits correspond to emotionality, warmth, desire and openness. In the context of social psychological intergroup research, it has been shown that the denial of human nature implies a form of mechanistic dehumanization. That is, members of social outgroups are commonly perceived rather in terms of automata than fully human. On the other hand, within the context of social robotics and the study of anthropomorphism, human nature represents an ideal measure to tap the humanization of nonhuman entities. To do so, we asked participants to what extent they would ascribe the following traits to the robot: nervous, curious, friendly, fun-loving, sociable, trusting, aggressive, distractible, impatient, and jealous. This 10-item measure yielded a reliable index ($\alpha = .91$).

Secondly, we measured the extent to which participants attributed mind to the robot. We did so by asking them to rate the robot with regard to 24 mental capacities (e.g., the capacity to feel pain, hunger, or to make plans) adapted from [28, see also 17]. The mind attribution index was highly reliable ($\alpha = .91$).

V. Independent Variables

A. Voice Samples and Pretesting

We collected voice samples from both male and female students at Bielefeld University for the main study. Specifically, participants were asked to state the neutral sentence ‘It is quarter to three’ in a neutral manner. Voice samples were recorded in a sound-proof silent cabin to assure noise-free recordings. Subsequently, all voice samples were synthesized in order to resemble of a robotic voice. In a pretest, a sample of university students rated the speakers (human vs. robotic; male vs. female) regarding warmth, competence, human-likeness vs. robot-likeness, and vocal femininity vs. masculinity. This was done to assure that the synthetic voice would be perceived as more robot-like than the human voice.

Analogously, the goal was to identify voice stimuli that would be classified correctly in terms of vocal femininity / masculinity. As a result of such pretesting, we obtained one male and one female human voice sample that only differed in terms of vocal femininity or masculinity respectively. The chosen samples were then synthesized, so that ultimately, gendered human-like and robot-like voice samples were ready for use as stimulus materials in the main study.

B. Gender of Robot Voice

Participants either watched a short video clip that featured the robot Flobi speaking in a masculine vs. a feminine voice to manipulate the alleged gender of the robot.

We assumed that when listening to a robot of their own gender, participants would anthropomorphize the robot more strongly because of activation of elicited agent knowledge.

C. Voice Type

Moreover, we manipulated voice type: That is, we varied whether participants watched a video clip that featured the robot speaking in a synthetic, robot-like vs. a humanlike voice. This was done to induce greater judgments of anthropomorphism in those participants who would listen to the sound sample featuring a human voice due to the fact that a human voice would, again, activate elicited agent knowledge.

VI. Results

To investigate effects of the factors gender of robot voice, voice type and participant gender on the dependent measures, we conducted a 2 (gender of robot voice: male vs. female) x 2 (voice type: robotic vs. human) x 2 (participant gender: male vs. female) multivariate analyses of variance (MANOVA). Results will be reported separately for the respective dependent variables.

A. Manipulation Check

The manipulation check showed that the experimental manipulation of voice type and robot gender via voice cues proved effective: First, participants evaluated the human voice sample as more human-like ($M = 5.18$, $SD = 1.61$) than the synthesized human voice ($M = 1.88$, $SD = 1.49$), $t(57) = 8.09$, $p < .001$. Second, participants in the male voice condition judged the robot voice as more masculine ($M = 6.10$, $SD = 1.17$) than in the female voice condition ($M = 2.64$, $SD = 1.66$), $t(57) = -9.32$, $p < .001$. These results are in line with findings from extensive pretesting of the natural and synthesized male and female voice samples (see section on Independent Variables).

B. Likeability

In line with our prediction, the robot with the human voice was rated more likeable ($M = 2.79$, $SD = 1.16$) than the robot with the synthetic voice ($M = 2.23$, $SD = 1.49$), $F(1, 50) = 6.98$, $p = .01$.

The gender of robot by gender of participant interaction merely approached significance, $F(1, 50) = 3.70$, $p = .06$. This pattern of results was further inspected by means of
independent samples t-tests: In tendency, female participants rated the female robot more likeable ($M = 3.00, SD = 1.58$) than the male robot ($M = 2.44, SD = 1.06$), $t(29) = 1.17, p = .25$. Male participants, on the other hand, showed the tendency to judge the same-sex robot as more likeable ($M = 2.82, SD = 1.53$) than the opposite-sex robot ($M = 2.00, SD = 0.76$), $t(25) = -1.75, p = .09$ (two-tailed). No other statistically significant results were obtained for likeability.

C. Psychological Closeness to the Robot

With regard to psychological closeness to the robot, the gender of robot voice x gender of participant interaction was statistically significant, $F(1, 50) = 5.56, p = .02$. This pattern was further inspected using two-tailed t-tests: These revealed that - in tendency - female participants felt more psychological closeness to the female robot ($M = 2.11, SD = 1.04$) than to the male robot ($M = 1.71, SD = 0.58$), $t(29) = 1.36, p = .19$.

Male participants, on the other hand, produced a stronger effect on psychological closeness to the robot. They felt significantly closer to the male robot ($M = 2.26, SD = 1.04$) than to the female counterpart ($M = 1.57, SD = 0.51$), $t(25) = -2.14, p = .04$. No other statistically significant results were obtained for psychological closeness to the robot.

D. Contact Intention

As with likeability and similarity ratings, the only effect that emerged for contact intentions was the gender of robot voice x gender of participant interaction, $F(1, 50) = 6.61, p = .02$. Again, the interaction pattern was inspected using two-tailed t-tests:

Female participants reported slightly, but not significantly higher contact intentions toward the female robot ($M = 3.30, SD = 1.67$) than toward the male prototype ($M = 2.72, SD = 0.71$), $t(29) = 1.31, p = .20$. Male participants showed the same pattern of means regarding the same-sex robot ($M = 3.41, SD = 1.44$) relative to the opposite-sex robot ($M = 2.52, SD = 0.80$), $t(25) = -1.96, p = .06$.

E. Anthropomorphism

Regarding judgments of anthropomorphism, we obtained neither statistically significant main effects, nor two-way interactions in the MANOVA.

Importantly, however, our results yielded three-way interactions of gender of robot voice, voice type, and participant gender, both for the attribution of human nature, $F(1, 50) = 4.89, p = .03$, and for mind attribution, $F(1, 50) = 5.09, p = .03$.

Separate analyses of variance (ANOVA) were computed for human and robotic voice types to inspect attributions of human nature as function of gender of robot voice and gender of participant.

When analyzing interaction effects of gender of robot voice and gender of participants for the human voice condition, we obtained only a marginally significant interaction of the two factors, $F(1,29) = 3.01, p = .09$. For robotic voice type, the two-way interaction failed to reach significance, $F(1, 25) = 2.25, p = .15$. No other statistically significant effects emerged.

To shed more light on the three-way interaction that was obtained for mind attribution, we likewise conducted separate ANOVAs for voice type: When analyzing interaction effects of gender of robot voice and gender of participants for the human voice condition, we obtained a significant interaction of the two factors, $F(1,29) = 9.00, p = .005$. This is illustrated in Fig. 2.

![Fig. 2. Mean ratings of Mind Attribution as a function of Gender of Participant and Gender of Robot in the human voice condition.](image)

The interaction did not turn out statistically significant in the robotic voice condition, $F < 1$.

VII. DISCUSSION AND CONCLUSION

The goal of the present research was to put Epley’s Three-Factor Theory of Anthropomorphism [7] to a further empirical test. Epley and colleagues have recently proposed three core determinants of anthropomorphic judgments about inanimate agents, such as robots. However, thus far, not all of them have been examined experimentally yet.

Recently, [17] and [18] have already investigated the role of situational and dispositional aspects of “effectance motivation” in predicting anthropomorphism and HRI acceptance. Furthermore, initial experimental evidence supports the notion that “sociality motivation” predicts anthropomorphism [16]. Only in a more indirect fashion has previous research investigated effects of the cognitive factor, elicited agent knowledge.

For instance, [14] and [15] have pointed to the processes that Epley and colleagues more explicitly spell out in their larger theoretical framework on anthropomorphism when addressing how ‘common ground’ between user and machine predicted communication patterns between human-robot interaction partners. Furthermore, previous work on effects of visual gender cues on social judgments about the robot Flobi [20] has shown that visual cues activated stereotypical agent knowledge. This elicited agent knowledge in turn biased evaluations of male vs. female robot prototypes.

Because previous work has focused strongly on vocal cues to manipulate a robot’s character [8-9, 14], we also relied on auditory gender cues in the present research. Crucially, however, we go beyond existing research on gender stereotyping and social robotics in that we tested not only effects of robot gender, but also of type of voice along with participant gender in the current research.

We did so to activate elicited agent knowledge, both at the level of gender (for the male vs. female subsamples,
respectively) as well as at the level of the human-likeness vs. robot-likeness of the voice. Accordingly, we predicted that if a human voice were implemented in a robot, this should elicit higher ratings on all assessed dependent variables including measures of HRI acceptance and anthropomorphism.

Interestingly, however, we only obtained a main effect of type of voice with regard to likeability ratings. Importantly, we found that a human voice resulted in higher estimated likeability of the system. With regard to psychological closeness, we found that particularly male participants experienced more psychological closeness to a same-sex robot than toward a robot of the opposite sex.

Female participants produced the same patterns of mean ratings, however, effects were less pronounced. Similar findings were obtained for contact intentions, even though in this case, the effect in the male subsample was only marginally significant. However, the results point to the notion that the activation of elicited agent knowledge results in stronger feelings of psychological closeness toward, and willingness for future contact with a technical system. Thus, by implementing features that strengthen the accessibility of such anthropocentric knowledge could be beneficial for HRI in general.

Three-way interactions were obtained on our measures of anthropomorphism, namely human nature and mind attribution. For mind attribution, effects of elicited agent knowledge were most evident in the human voice condition, as in this condition, the gender of robot voice x gender of participant interaction turned out significant. In the robotic voice condition, this was not the case. Similar patterns were obtained for human nature attribution, even though these were less pronounced and turned out only marginally significant in the human voice condition.

Taken together, the present results fit nicely with Epley’s notion of elicited agent knowledge activation, even though not all experimental hypotheses, particularly on main effects of type of voice were substantiated empirically.

However, just as in [19], where participants attributed more mind and other typically human traits to a robot that ostensibly belonged to their national ingroup, participants in the current research also appeared to favor the same-sex robot in their ratings of mind attribution, specifically when the robot used human rather than synthetic speech.

Clearly, future research should substantiate these initial findings further, because the current results bear practical significance for applied settings. For instance, we have found that participants seem to attribute more likeability to a technical system that used human speech, moreover, they tended to experience more psychological and showed more interest in contact. Finally, they even attributed more mind to a system that resembled themselves even though we have used only very subtle auditory cues. These results imply a social projection mechanism that may help facilitate HRI by making it possible to experience psychological common ground with a technical system that elicits anthropocentric knowledge.

Follow-up research should address the additive effect of visual and vocal gender cues on social perception and on “discriminatory” behaviors toward gendered robots. This would show that the activation of elicited agent knowledge not only leads to differential social judgments about a robot, but that it would also bias the behavior we would demonstrate toward it. Therefore, it would ultimately be desirable to examine effects of robot configuration in applied settings, e.g. in care-centers for the elderly. In such contexts, potential end-users of robot assistants can provide the ultimate answer to the question as to which factors are indeed key in determining HRI acceptance and anthropomorphic inferences.

**ACKNOWLEDGMENT**

The authors thank Simon Bobinger for his valuable contributions to this research. This research was funded as part of the Cluster of Excellence in Cognitive Interaction Technology (CoE 277) by the German Research Council (DFG).

**REFERENCES**


