“Ick bin een Berlina”: dialect proficiency impacts a robot’s trustworthiness and competence evaluation

Katharina Kühne1*, Erika Herbold1, Oliver Bendel2, Yuefang Zhou1 and Martin H. Fischer1

1Division of Cognitive Sciences, University of Potsdam, Potsdam, Germany, 2School of Business FHNW, Brugg-Windisch, Brugg, Switzerland

Background: Robots are increasingly used as interaction partners with humans. Social robots are designed to follow expected behavioral norms when engaging with humans and are available with different voices and even accents. Some studies suggest that people prefer robots to speak in the user’s dialect, while others indicate a preference for different dialects.

Methods: Our study examined the impact of the Berlin dialect on perceived trustworthiness and competence of a robot. One hundred and twenty German native speakers (M_age = 32 years, SD = 12 years) watched an online video featuring a NAO robot speaking either in the Berlin dialect or standard German and assessed its trustworthiness and competence.

Results: We found a positive relationship between participants’ self-reported Berlin dialect proficiency and trustworthiness in the dialect-speaking robot. Only when controlled for demographic factors, there was a positive association between participants’ dialect proficiency, dialect performance and their assessment of robot’s competence for the standard German-speaking robot. Participants’ age, gender, length of residency in Berlin, and device used to respond also influenced assessments. Finally, the robot’s competence positively predicted its trustworthiness.

Discussion: Our results inform the design of social robots and emphasize the importance of device control in online experiments.

KEYWORDS
competence, dialect, human-robot interaction, robot voice, social robot, trust

1 Introduction

1.1 Factors influencing robot’s acceptance

Social robots are becoming more common in various social aspects of human life, such as providing interpersonal care, tutoring, and companionship (Belpaeme et al., 2018; Bendel, 2021; Breazeal, 2017; Broadbent, 2017; Zhou and Fischer, 2019; for review, see e.g., Cifuentes et al., 2020; Woo et al., 2021; Henschel et al., 2021). Unlike most manufacturing or surgical robots, a social robot is designed to have a physical body and interact with humans in a way that aligns with human behavioral expectations.
Specifically, a humanoid robot is a type of a social robot with a body shape resembling a human, including a head, two arms, and two legs (Broadbent, 2017). According to Bendel (2021), social robots are sensorimotor machines created to interact with humans or animals. They can be identified through five key aspects. These are non-verbal interaction with living beings, verbal communication with living beings, representation of (aspects of or features of) living beings (e.g., they have an animaloid or a humanoid appearance or natural language abilities), proximity to living beings, and their utility or benefit for living beings. The assumption is that an entity is a social robot if four of these five dimensions are met. It can be hypothesized that the ability to speak and the voice used are likely to be among the central features of social robots. The present study focused on the role of speech to better understand social interactions with robots.

Which factors affect whether a person accepts a robot as a social interaction partner? Some of these factors include human-related aspects such as previous exposure to robots, the age and gender of the person interacting with robots (Broadbent et al., 2009; Kuo et al., 2009; Nomura, 2017; but also see Bishop et al., 2019; for a review, see Naneva et al., 2020). While it is generally observed that increased exposure to social robots corresponds to more favorable attitudes toward them, the evidence regarding age and gender as factors influencing acceptance is inconclusive. Previous studies suggested that older individuals and females tend to have less positive attitudes toward robots (Kuo et al., 2009; May et al., 2017). However, a systematic review (Naneva et al., 2020) contradicted this conclusion. According to this analysis, age and gender do not appear to have a significant impact on acceptance of social robots. Additionally, personality features might also play a role. According to Naneva et al. (2020), there is a positive correlation between acceptance of robots and the personality traits of agreeableness, extroversion, and openness, while conscientiousness and neuroticism do not appear to have any significant impact (Esterwood et al., 2021).

Apart from some human-related factors discussed above that could impact robot acceptance, many other factors that potentially influence human-robot interaction outcome concern the robot itself, including the purpose it is used for and its appearance. Whereas multiple studies demonstrated that users prefer human-like robots (Esposito et al., 2019; 2020), the systematic review by Naneva et al. (2020) could not find clear evidence for that. Here, we focus on some robot-related factors, in particular its voice, to motivate a novel research question, as will be reviewed in the next few paragraphs.

1.2 Anthropomorphism in robot design and its impact on interaction

People tend to ascribe human traits to non-human entities. There are two aspects to consider. Firstly, users attribute certain human behaviors to the robot by projecting their own expectations onto it. Secondly, individuals intentionally program the robot with human behaviors. Companies provide robots with a variety of physical appearances and voices that differ in gender, age, accent, and emotional expression, to cater to a wide range of needs and preferences of their users (Epley et al., 2007). An anthropomorphic robot design enables a more natural interaction with robots because people can rely on behaviors familiar from human-human interactions (Codic et al., 2017). Moreover, a humanoid appearance results in more positive evaluation of the robot (Biermann et al., 2020).

1.3 Robot’s voice in trust and competence evaluation

To have a productive interaction, humans need to have confidence in and trust a social robot (Marble et al., 2004). Trust can influence the success of human-robot collaboration and determine future robot use (Freedy et al., 2007). In human-robot interactions, trust has been the subject of extensive research (Dunning and Fetchenhaus, 2011). Crucially, multiple studies have indicated that trust does not necessarily result from a logical evaluation of the probabilities of different outcomes and benefits involved in a given situation. Rather, it seems to stem from non-rational factors, such as feelings and emotions. Factors that contribute to trust are linked to the attributes of both the person, the circumstances, and their interplay (Evans and Krueger, 2009; for review see Thielmann and Hilbig, 2015). In particular, being part of the same group can heighten trust levels (Evans and Krueger, 2009).

Trust in human-robot interaction is defined as “the attitude that an agent will help achieve an individual's goals in a situation characterized by uncertainty and vulnerability” (Lee and Sec, 2004) or as “the reliance by an agent that actions prejudicial to their wellbeing will not be undertaken by influential others” (Hancock et al., 2011). These definitions imply that humans who trust a robot believe that it will not harm them or can be relied on in fulfilling tasks (Law and Scheutz, 2021).

Although numerous factors can impact trust in artificial agents (as demonstrated by Schaefer et al., 2016; Hancock et al., 2011 in their respective meta-analyses; for systematic review see Rheu et al., 2021; Law and Scheutz, 2021), the voice of a robot is considered one of the most critical factors in determining trust specifically related to robots.

In a questionnaire study conducted by Dautenhahn et al. (2005), most of the respondents expressed a desire for a robotic companion that can communicate in a way that is very similar to a human. Individuals also tended to get closer to a robot that had a human-like voice, in contrast to a robot with an artificially synthesized voice (Walters et al., 2008). Human-like voices were perceived as less uncanny and rated higher in terms of qualities such as sympathy, credibility, and trustworthiness (K. Kühne et al., 2020). Robots that had human-like voices were considered to be more efficient and were remembered more easily (Rodero, 2017). Finally, artificial agents with a human-like voice were perceived as more competent and credible (Sims et al., 2009; Fischer, 2021; Kim et al., 2022).

Competence is another attribute that is often intuitively assessed in everyday interactions (Kovarsky et al., 2013; Abele et al., 2021). The Behavioral Regulation Model defines confidence as the likelihood of task achievement (Ellemers et al., 2013). Alongside warmth, confidence underlies social evaluation and relies on such features as power, status, and resources (Rosenberg et al., 1968). In human–robot interaction, competence was one of the most important predictors of human preferences between different robot behaviors (Oliveira et al., 2019; Scheunemann et al., 2020). Also in
Evidence of the influence of dialect on the trust or competence of a robot is mixed. In general, according to the similarity-attraction theory, individuals tend to prefer artificial agents similar to themselves, for example, in terms of personality (Nass and Lee, 2000). However, similarity on a more superficial level, such as gender, was not found to predict trust (You and Robert, 2018).

In addition to identifying the speaker as a member of a particular geographical or national group, a dialect can also elicit favorable or unfavorable connotations and shape opinions about the speaker irrespective of the own group (H. Bishop et al., 2005). Listeners are sensitive to sociolinguistic information conveyed by a dialect or an accent. The standard language is typically viewed as prestigious and reliable, whereas regional accents tend to be regarded more unfavorably (H. Bishop et al., 2005; Tamagawa et al., 2011), so-called “accentism” (Foster and Stuart-Smith, 2023). However, certain languages may also have esteemed regional variations or dialects (H. Bishop et al., 2005).

Prejudices against dialects and their speakers cannot be ignored, as evaluations of dialects are often associated with evaluations of the corresponding population (Wiese, 2012). A meta-analysis by Fuertes et al. (2012) revealed that a spoken dialect is perceived as a sign of lower intelligence and social class. According to Wiese (2012), individuals who do not use the standard language are often viewed as linguistically incompetent. Furthermore, Fuertes et al. (2012) found that a spoken dialect can lower the perception of competence in general.

There are conflicting findings regarding the effects of different dialects on the perception of robots. On the one hand, imparting the standard language to a robot was shown to increase its trustworthiness and competence (Torre and Maguer, 2020). As an example, only around 4% of Torre and Maguer’s (2020) participants wanted the robot to have the same accent as they had, whereas 37% preferred a robot speaking the Standard Southern British English. Similar findings were obtained by Andrist et al. (2015): More native Arabic speakers complied with the robots who were speaking standard Arabic. For the dialect-speaking robot, the compliance depended on other factors. Namely, robots speaking with both high knowledge and high rhetorical ability were complied with more. Another study found that a synthetic agent with Austrian standard accent was perceived as possessing higher levels of education, trustworthiness, competence, politeness, and seriousness (Krenn et al., 2017).

On the other hand, robots speaking a dialect, in this case, Franconian, were rated as more competent (Lugrin et al., 2020). Unlike in Torre and Maguer (2020), the evaluation of competence depended on the participants’ own performance in the dialect. Those who spoke in dialect more frequently rated the dialect-speaking robot as more competent. In contrast to that, V. Kühne et al. (2013) found that participants liked a dialect-speaking robot more, irrespective of their own dialect performance. In the same vein, a robot was accepted in Norwegian hospitals more when it spoke the Trøndersk dialect (Sørnes and Fostervold, 2021). This preference could have been impacted by the comfortable and pleasant connotation conveyed by the Trøndersk dialect. To embrace these discrepancies, there is currently an ongoing project to develop an optimal language or accent for an artificial agent to speak (Foster and Stuart-Smith, 2023).

1.5 The impact of dialect-related social classifications and group identity

Interestingly, dialect-related social classifications and the sense of being part of a group based on accent or dialect are more robust than those resulting from gender or ethnicity (Kinzler et al., 2010). A dialect or accent refers to how individuals from diverse regions or social groups articulate words and phrases, leading to differences in their accent and speech patterns. While dialects and accents are interconnected, they are not identical. Dialects encompass a wider range of linguistic aspects, including vocabulary, grammar, and sentence structure, whereas accents primarily involve differences in pronunciation (Sikorski, 2005; for more detailed information on the topic of accent and dialect, see Planchenault and Poljak, 2021).
In summary, standard language-speaking robots were perceived as more trustworthy or likable presumably due to the in-group bias and accentism, while according to other studies, participants preferred robots that spoke with a dialect. However, the preference for dialect-speaking robots was often influenced by human-related factors, namely, the participants’ proficiency or performance in that dialect (Lugrin et al., 2020).

Most of the research on the utilization of dialect in robots has been conducted in Anglo-Saxon countries (Früh and Gasser, 2018). As for German-speaking countries, V. Kühne et al. (2013) found that a Rhine-Ruhr dialect-speaking virtual robot was perceived as more likable. Another study by Früh and Gasser (2018) also reports more positive attitudes toward a dialect-speaking care robot Lio in Switzerland. Importantly, in Switzerland, a dialect serves strongly as a means of social demarcation. However, a most recent study with a service robot Pepper in a hotel context showed that using the local dialect did not affect robot acceptance and attitudes (Steinhaeusser et al., 2022). The study was conducted online and participants speaking the Flanconian dialect vs. standard German were randomly assigned to the dialect or standard language conditions. While there was a non-significant tendency for individuals who spoke a dialect to have a more negative attitude toward a robot that used that same dialect, this could potentially be attributed to the use of Pepper’s text-to-speech plugin to synthesize the dialect and accent. People with a local accent may have been more likely to notice any mistakes or errors in the robot’s synthesized speech, which could in turn have influenced their attitudes towards it.

To address the inconsistencies reviewed above, we conducted an online study among Berlin and Brandenburg residents in order to investigate the relationship between the participants’ proficiency and performance in the Berlin dialect and their trust in a robot, and the robot’s competence evaluation.

1.6 The present study

From 1500 onwards, the Berlin dialect emerged as a unique local language variety, replacing Low German in the region. The Berlin dialect is associated with the working class and often portrayed as a proletarian language by media figures who depict it as a dialect spoken by simple, but likable people. Additionally, the Berlin dialect is intentionally employed as a stylistic choice to establish a sense of closeness with a specific audience, as observed in its written representation in daily newspapers (Wiese, 2012). Specific features of the Berlin dialect can be found in Stickel (1997).

Dialect proficiency means the self-evaluated ability to speak the dialect, whereas dialect performance denotes the frequency with which the participants speak the dialect. We formulated six hypotheses for our study. The first hypothesis was that the standard German-speaking robot would be trusted more and evaluated as more competent than the dialect-speaking robot (H1). The next two hypotheses posited that participants with (H2) higher dialect proficiency and (H3) higher dialect performance would trust the robot more than those with lower dialect proficiency and performance. The fourth and fifth hypotheses were that participants with (H4) higher dialect proficiency and (H5) higher dialect performance would evaluate the robot’s competence higher than those with lower dialect proficiency and performance. Finally, we expected that the robot’s competence would predict the trust ratings (H6). Hypotheses H2—H6 were tested independently for the dialect-speaking robot (H2a, H3a, H4a, H5a, H6a) and the standard German-speaking robot (H2b, H3b, H4b, H5b, H6b) as alternatives. We tested these formulated hypotheses in an online experiment with German-speaking participants using the NAO robot.

2 Materials and methods

2.1 Participants and procedure

The experiment was programmed and run using the online Gorilla Experiment Builder research platform (Anwyl-Irvine et al., 2020) and lasted approximately 30 min. The participants were recruited via the subject pool system SONA at the University of Potsdam. All the participants submitted their informed consent at the beginning of the experiment by clicking the corresponding checkbox and were reimbursed with course credits for their participation. They were instructed to first watch a video and then answer the survey questions honestly and spontaneously. The type of the video (Berlin dialect or Standard German) was counterbalanced between participants. After the survey, the participants were asked to fill in a demographic questionnaire, including questions about their age, gender, native language, dialect proficiency, dialect performance, and duration of residence in Berlin. Finally, participants were debriefed and given a link to enter their internal subject pool ID for receiving a credit.

The study was conducted in accordance with the guidelines laid down in the Declaration of Helsinki and in compliance with the ethics policy of the University of Potsdam. No explicit approval was needed because the methods were standard. There were no known risks and participants gave their informed consent. The study and the procedure were already evaluated by professional psychologists to be consistent with the ethical standards of the German Research Foundation, including written informed consent and confidentiality of data as well as personal conduct.

An a priori power analysis was conducted using G*Power (Faul et al., 2007) to determine the minimum sample size required to test the study hypothesis. Results indicated the required sample size to achieve 80% power for detecting a medium effect, at a significance criterion of α = .05, was N = 68 per robot group for linear regression with two predictors (N = 136 in total).

2.2 Stimuli materials

We used a video lasting 31 s, showcasing the humanoid robot NAO (Aldebaran—SAS)1. In the video, the robot was positioned on a table and was in motion while providing details about a painting situated in the top right portion of the wall. The painting was pixelated to avoid copyright infringement. A snapshot from the video is depicted in Figure 1.

1 The video material used was made and provided by Tristan Kornher, student at the University of Potsdam.
The robot in the video used a male human voice to speak. The speech was recorded twice by the same speaker—once in standard German and once in the Berlin dialect. The transcription can be found in Supplementary Materials.

We opted to use a human voice based on earlier studies, which indicated that people prefer less robotic-sounding voices as they feel more at ease while listening to them (Dong et al., 2020; Kühne et al., 2020). Natural human voices are generally perceived as more trustworthy and competent compared to synthetic voices (Craig and Schroeder, 2017; Kühne et al., 2020; Sims et al., 2009). Moreover, listening to a synthetic voice can increase one’s cognitive load (Francis and Nusbaum, 2009; Simantiraki et al., 2018) which, in its turn, can lead to trust misplacement (Duffy and Smith, 2014).

We selected a male voice because research suggests that NAO is more commonly associated with a male voice (Behrens et al., 2018). The stimuli can be found at: https://osf.io/pfgg6/.

2.3 Measures

2.3.1 Independent variables

2.3.1.1 Demographic factors
The following demographic factors were measured: age, gender, native language, and duration of residence in Berlin (in years).

2.3.1.2 Dialect proficiency
The dialect proficiency was measured using a single item: “How well can you speak the Berlin dialect?”. The answers were given on a seven-point Likert scale from 1 (Not at all) to 7 (Very well).

2.3.1.3 Dialect performance
The dialect performance was measured using a single item: “In everyday life, I usually speak the Berlin dialect”. The answers were given on a seven-point Likert scale from 1 (Does not apply at all) to 7 (Applies totally).

2.3.1.4 Device type
Device type was automatically measured by the experiment system as “mobile,” “tablet,” or “computer”.

2.3.2 Dependent variables

2.3.2.1 Trust
We used the Scale of Trust in Automated Systems (Jian et al., 2000) to access the level of trust participants had toward the robot featured in the video. The scale consists of 12 items, measured on a seven-point Likert scale from 1 (Do not agree at all) to 7 (Fully agree), and was specifically designed to measure trust towards automated systems, such as robots. To suit the study’s German setting, the items were translated into German, and the word “system” in each item was replaced with “robot” to better relate to the robot shown in the video. Sample items were: “I can trust the robot” (“Ich kann dem Roboter vertrauen”); “The robot is dependable” (“Der Roboter ist verlässlich”). Supplementary Table S1 displays the original items and their corresponding German translations. Additionally, an extra attention-testing item was added to the scale, which instructed participants to choose response option 7 (Fully agree) as their response.

2.3.2.2 Competence
We used the Robotic Social Attribute Scale (RoSAS) (Carpinella et al., 2017) to measure the competence evaluation of the featured robot. The scale consists of 6 items, measured on a seven-point Likert scale from 1 (Do not agree at all) to 7 (Fully agree). Sample items were: “The robot is interactive” (“Der Roboter ist interaktiv”); “The robot is knowledgeable” (“Der Roboter ist sachkundig”). Supplementary Table S2 displays the original items and their corresponding German translations. Additionally, an extra attention-testing item was added to the scale, which instructed participants to choose the response option 1 (“Do not agree at all”) as their response.
2.4 Sample and data pre-processing

One hundred and thirty-seven participants (94 females, 41 males, 2 non-binary), Mean age = 33 years, SD = 14 years) took part in the experiment. Eight participants were excluded from the analysis because their native language was not German. Nine participants were further excluded from the analysis because they failed the attention test items in both scales. This yielded the ultimate sample size N = 120 (Mean age = 32 years, SD = 12 years; 81 female, 38 male, 1 non-binary). Additionally, data from the trust items of two participants and data from the competence items of three participants were excluded because they failed the attention test items in the respective scale. The remaining data of these five participants was still used.

Data preparation and analyses were done using Microsoft® Excel® for Microsoft 365 and SPSS Version v.29 software package. Figures were built in R (R Core Team, 2020). The normality of the data distribution was confirmed using a Kolomogorov-Smirnov test. Before conducting the multiple regression analysis, the distributional assumptions for the multiple regression were assessed. The regression analysis treated the gender category of “non-binary” as missing data.

3 Analysis and results

3.1 Trust

First, we employed a two-tailed independent samples t-test to examine the level of trust between the dialect-speaking robot and the standard German-speaking robot in all participants. Even though there was a minor trend in favor of trusting the standard German-speaking robot more (M = 4.716, SD = 1.259) than the dialect-speaking one (M = 4.591, SD = 1.056), this difference was not statistically significant (t (116) = −0.583, p = .561). Thus, we failed to confirm H1a. Participants did not trust the standard German-speaking robot significantly more than the dialect-speaking robot.

To examine if participants with higher dialect proficiency would trust the dialect-speaking robot more than those with lower dialect proficiency, we conducted a multiple regression analysis, using the enter method. In the first step, we added only dialect proficiency as predictor. In the second step, we added control variables: age, gender, duration of residence in Berlin, and device type. In line with the H2a hypothesis, only dialect proficiency explained a significant amount of the variance in the value of trust in the dialect-speaking robot (β = −.017). However, age, gender, duration of residence in Berlin, and device type were significant predictors of trust. The standard German-speaking robot was more trusted by individuals who were older, female, had a shorter duration of residence in Berlin, and used a computer device for watching the experimental videos.

Finally, we conducted another multiple regression to examine if participants with higher dialect performance would trust the standard German-speaking robot more than those with lower dialect performance. In the first step, we added only dialect performance as predictor. In the second step, we added control variables: age, gender, duration of residence in Berlin, and device type. Contrary to the H3a hypothesis, dialect performance was not a significant predictor of trust in the dialect-speaking robot (β = .086, t (53) = .628, p = .533, F (1, 53) = 0.394, R² = .007, R² Adjusted = −.011). However, age, gender, duration of residence in Berlin, and device type were significant predictors of trust. The standard German-speaking robot was more trusted by individuals who were older, female, had a shorter duration of residence in Berlin, and used a computer device for watching the experimental videos.

In summary, for the standard German-speaking robot, age, gender, duration of residence in Berlin, and device type were significant predictors of trust, when together in model with dialect proficiency. We found no evidence for H2b and H3b.

The results are summarized in Table 1 and Table 2. Figure 2 presents a visual summary of the outcomes obtained from regression analyses that assessed how dialect proficiency predicted trust in both the standard German-speaking and dialect-speaking robot.

3.2 Competence

Again, we used a two-tailed independent samples t-test to examine the level of competence between the dialect-speaking robot and the standard German-speaking robot in all participants. The findings were similar for the evaluation of trust. While there was a descriptive tendency to rate the standard German-speaking robot as more competent (M = 3.831, SD = 0.947) than the dialect-speaking robot (M = 3.777, SD = 0.999), the difference was not statistically significant (t (115) = −0.303, p = .763). Thus, we failed to confirm H1b. Participants did not evaluate the standard German-speaking
Further, to examine if participants with higher dialect proficiency would evaluate the standard German-speaking robot as more competent than those with lower dialect proficiency, we conducted a multiple regression using the enter method. In the first step, we added only dialect proficiency as predictor. In the second step, we added control variables: age, gender, duration of residence in Berlin, and device type. Contrary to the H4b hypothesis, dialect proficiency alone was not a significant predictor of competence in the standard German-speaking robot (β = .086, t (52) = 0.536, p = .561, F (1, 52) = 0.389, R² = .007, R² Adjusted = −.012). However, when controlled for age, gender, duration of residence in Berlin, and device type, it did explain a reliable amount of variance in the value of competence, together with duration of residence in Berlin (β = .695, t (48) = 3.463, and β = −.824, t (48) = −3.735, respectively, p < .001, F (5, 48) = 4.634, R² = .326, R² Adjusted = .255). Age, gender, and device type did not contribute to the final model.

To see if participants with higher dialect performance would evaluate the standard German-speaking robot as more competent than those with lower dialect performance, we conducted a multiple regression using the enter method. In the first step, we added only dialect performance as predictor. In the second step, we added control variables: age, gender, duration of residence in Berlin, and device type. Contrary to the hypothesis H5b, dialect performance alone was not a significant predictor of competence in the standard German-speaking robot (β = .051, t (52) = 0.365, p = .717, F (1, 52) = 0.133, R² = .003, R² Adjusted = −.017). However, when controlled for age, gender, duration of residence in Berlin, and device type, it did explain a

### TABLE 1 Results of the Regression Analysis on the Outcome Variable Trust with Dialect Proficiency as Predictor.

<table>
<thead>
<tr>
<th>Model</th>
<th>Dialect-speaking robot</th>
<th>Standard German-speaking robot</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>β</td>
<td>SE</td>
</tr>
<tr>
<td>1</td>
<td>Constant</td>
<td>0.227</td>
</tr>
<tr>
<td></td>
<td>Proficiency</td>
<td>.272</td>
</tr>
<tr>
<td>R²</td>
<td>.074</td>
<td>.009</td>
</tr>
<tr>
<td>R² Adjusted</td>
<td>.059</td>
<td>.009</td>
</tr>
<tr>
<td>p</td>
<td>&lt;.05</td>
<td>&lt;.05</td>
</tr>
<tr>
<td>2</td>
<td>Constant</td>
<td>0.532</td>
</tr>
<tr>
<td></td>
<td>Proficiency</td>
<td>.345</td>
</tr>
<tr>
<td></td>
<td>Age</td>
<td>.174</td>
</tr>
<tr>
<td></td>
<td>Gender</td>
<td>−.144</td>
</tr>
<tr>
<td></td>
<td>Duration</td>
<td>−.179</td>
</tr>
<tr>
<td></td>
<td>Device</td>
<td>.082</td>
</tr>
<tr>
<td>R²</td>
<td>.119</td>
<td>.001</td>
</tr>
<tr>
<td>R² Adjusted</td>
<td>.040</td>
<td>.001</td>
</tr>
<tr>
<td>p</td>
<td>&lt;.05</td>
<td>&lt;.05</td>
</tr>
</tbody>
</table>

Note: Dialect-speaking robot N = 63. Standard-speaking robot N = 57.
Method: enter. Significant results are marked in bold.
TABLE 2 Results of the Regression Analysis on the Outcome Variable Trust with Dialect Performance as Predictor.

<table>
<thead>
<tr>
<th>Model</th>
<th></th>
<th>Dialect-speaking robot</th>
<th>Standard German-speaking robot</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>β</td>
<td>SE</td>
<td>t</td>
</tr>
<tr>
<td>1</td>
<td>Constant</td>
<td>0.206</td>
<td>21.069</td>
</tr>
<tr>
<td></td>
<td>Performance</td>
<td>0.208</td>
<td>0.090</td>
</tr>
<tr>
<td></td>
<td>R²</td>
<td>.043</td>
<td></td>
</tr>
<tr>
<td></td>
<td>R²Adjusted</td>
<td>.027</td>
<td></td>
</tr>
<tr>
<td></td>
<td>p</td>
<td>.105</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>Constant</td>
<td>0.544</td>
<td>7.742</td>
</tr>
<tr>
<td></td>
<td>Performance</td>
<td>.142</td>
<td>0.108</td>
</tr>
<tr>
<td></td>
<td>Age</td>
<td>0.174</td>
<td>0.014</td>
</tr>
<tr>
<td></td>
<td>Gender</td>
<td>−.129</td>
<td>0.313</td>
</tr>
<tr>
<td></td>
<td>Duration</td>
<td>.004</td>
<td>0.064</td>
</tr>
<tr>
<td></td>
<td>Device</td>
<td>0.074</td>
<td>0.271</td>
</tr>
<tr>
<td></td>
<td>R²</td>
<td>.081</td>
<td></td>
</tr>
<tr>
<td></td>
<td>R²Adjusted</td>
<td>−.001</td>
<td></td>
</tr>
<tr>
<td></td>
<td>p</td>
<td>.434</td>
<td></td>
</tr>
</tbody>
</table>

Note: Dialect-speaking robot N = 63. Standard-speaking robot N = 57.
Significant results are marked in bold.

reliable amount of variance in the value of competence, together with duration of residence in Berlin and device type (β = .410, t (48) = 2.433; β = −.529, t (48) = −2.768; and β = .281, t (48) = 2.188 respectively, p < .05, F (5, 48) = 3.193, R² = .250, R²Adjusted = .171). age and gender did not contribute to the final model.

In summary, for the standard German-speaking robot both dialect proficiency and dialect performance were significant predictors of competence, but only when controlled for age, gender, duration of residence in Berlin, and device type. Hypotheses H4b and H5b could be partially confirmed. Duration of residence in Berlin and device type were also reliable predictors of competence for the standard German-speaking robot.

The results are summarized in Table 3 and Table 4.

Figure 3 presents a visual summary of the outcomes obtained from regression analyses that assessed how dialect proficiency predicted competence in both the standard German-speaking and dialect-speaking robot.

3.3 Association between robot’s competence and trust

Lastly, we sought to determine if the evaluation of a robot’s competence could predict the degree of trust that was placed in the robot. Indeed, for both the dialect-speaking robot (β = .631, t (59) = 6.249, F (1, 59) = 39.049, p < .001, R² = .398, R²Adjusted = .388) and the standard German-speaking robot (β = .646, t (52) = 6.096, F (1, 52) = 37.164, p < .001, R² = .417, R²Adjusted = .406), competence was a significant predictor of trust. Both H6a and H6b could be confirmed. Figure 4 presents a visual representation of the outcomes of the regression analyses.

The data set and the analysis script can be found at: https://osf.io/pfgg6/.

4 Discussion

4.1 Proficiency and performance in the Berlin dialect and evaluation of competence and trust

Our study investigated verbal aspects of human robot interaction quality. Specifically, we examined the association between participants’ proficiency and performance in the Berlin dialect and their evaluation of competence and trust in a NAO robot that spoke either with or without this dialect. The study was conducted online, and dialect proficiency was defined as the self-evaluated ability to speak the Berlin dialect, while dialect performance referred to the frequency of dialect used by the participants.

In general, although the difference in trust and competence ratings were not significant, our findings tend to be consistent
FIGURE 2
Regression Analysis for Dialect Proficiency as a Predictor of Trust in the Standard German-speaking and the Dialect-speaking Robot

Note: The orange solid line represents the regression slope for the dialect-speaking robot. The dark blue long dashed line represents the regression slope for the standard German-speaking robot.

TABLE 3 Results of the Regression Analysis on the Outcome Variable Competence with Dialect Proficiency as Predictor.

<table>
<thead>
<tr>
<th>Model</th>
<th>Dialect-speaking robot</th>
<th>Standard German-speaking robot</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>β</td>
<td>SE</td>
</tr>
<tr>
<td>1</td>
<td>Constant</td>
<td>0.226</td>
</tr>
<tr>
<td></td>
<td>Proficiency</td>
<td>0.047</td>
</tr>
<tr>
<td></td>
<td>R²</td>
<td>.002</td>
</tr>
<tr>
<td></td>
<td>R² Adjusted</td>
<td>−.014</td>
</tr>
<tr>
<td></td>
<td>p</td>
<td>.718</td>
</tr>
<tr>
<td>2</td>
<td>Constant</td>
<td>0.502</td>
</tr>
<tr>
<td></td>
<td>Proficiency</td>
<td>0.293</td>
</tr>
<tr>
<td></td>
<td>Age</td>
<td>−.216</td>
</tr>
<tr>
<td></td>
<td>Gender</td>
<td>−.152</td>
</tr>
<tr>
<td></td>
<td>Duration</td>
<td>−.222</td>
</tr>
<tr>
<td></td>
<td>Device</td>
<td>0.079</td>
</tr>
<tr>
<td></td>
<td>R²</td>
<td>.121</td>
</tr>
<tr>
<td></td>
<td>R² Adjusted</td>
<td>.042</td>
</tr>
<tr>
<td></td>
<td>p</td>
<td>.193</td>
</tr>
</tbody>
</table>

Note: Dialect-speaking robot N = 63. Standard-speaking robot N = 57.
Significant results are marked in bold.

with previous studies conducted by Torre and Maguer (2020) and Andrist et al. (2015) which also found that people preferred a robot that speaks in standard language. This is in line with the overall research suggesting that individuals who speak the standard language are perceived as more competent (Fuertes et al., 2012). However, our findings are contradictory to the results of V. Kühne et al. (2013) and Früh and Gasser (2018) where a robot speaking in dialect was viewed more positively. It is essential
that their experiments were conducted in Switzerland, as the local dialect plays a crucial role, in distinguishing insiders from outsiders. Further, similar to Lugrin et al. (2020) we demonstrated that participants’ ratings of the robot’s trust and competence were influenced by their own proficiency in the dialect, but our study provided more nuanced results.

Importantly, as expected, the competence of the robot significantly predicted trust. Namely, the more competent the robot...
was rated by the participants, the more they trusted it. This is in line with previous research (Hancock et al., 2011; Kraus et al., 2018; Steain et al., 2019; Christoforakos et al., 2021). Competence is perceived as an ability to carry out behavioral intentions (Kulms and Kopp, 2018). Being a positive quality, it creates a more favorable impression of the trustee. As a major dimension of social cognition postulated by the Stereotype Content Model, competence has been observed to foster the establishment of trust in interactions between humans (Fiske et al., 2007). Also according to another model, competence and benevolence of the trustee are positively related to trust (Mayer et al., 1995). Thus, we report evidence indicating that social mechanisms observed in human-human interactions can be transferred to human-robot interactions.

In the following paragraphs we will discuss the findings in detail. In the first place, although there was a slight trend of higher trust and competence evaluation for the standard German-speaking compared to the dialect-speaking robot for all participants, the difference was not statistically significant. The standard German-speaking robot and the dialect-speaking robot received largely comparable ratings in terms of both competence and trustworthiness.

Nevertheless, there were systematic differences in ratings between the two robots. Consider first the ratings obtained for the dialect-speaking robot. For the dialect-speaking robot, only dialect proficiency was a significant predictor of trust, with individuals who considered themselves more proficient in speaking the Berlin dialect having higher levels of trust. The other predictors (dialect performance, age, gender, duration of residence, and device type) did not have a significant contribution to the final statistical model of the ratings on trust. Our analysis for the outcome variable competence showed no significant predictors. Dialect proficiency, dialect performance, age, gender, duration of residence, and device type did not significantly contribute to the final model of participants' rating. Thus, for the dialect-speaking robot, only one reliable association was found, namely, that between dialect proficiency and the trust in robots. The more proficient the participants were in the Berlin dialect, the more they trusted the dialect-speaking NAO, exactly in the sense of the similarity-attraction theory (Nass and Lee, 2000). None of the factors were found to be predictive of the level of robot's competence.

For the standard German-speaking robot, the findings were more complex. We found that the final model included age, gender, duration of residence, and device type as significant predictors of trust, but only when included into the model together with dialect proficiency. Individuals who were older, female, had a shorter duration of residence in Berlin, and used a computer device for watching the experimental videos were found to trust the standard German-speaking robot more. Dialect performance did not make a significant contribution to the model.

Finally, dialect proficiency, dialect performance, duration of residence, and device type were significant predictors of competence, indicating that those who were more proficient in speaking the Berlin dialect, spoke it more often, had a shorter duration of residence in Berlin, and used a computer device for watching the experimental videos found the standard German-speaking robot more competent.

For the standard German-speaking robot, general factors such as age and gender appeared to be predictive of the trust level, while the participants' dialect proficiency and performance only played a role in the evaluation of competence. This finding collaborates with earlier research reporting the importance of demographic factors on robot's perception (Naneva et al., 2020). Similar to results obtained by K. Kühne et al. (2020), female participants evaluated the robot as more trustworthy. In comparison to that research, however, we found that, as participants’ age increased, their trust in the standard German-speaking robot also increased.

In conclusion, again following the principles of the similarity-attraction theory (Nass and Lee, 2000), participants who had been living in Berlin for a shorter period, presumably were less likely
to be influenced by the Berlin dialect, were more likely to trust the robot that spoke in standard German and found it more competent.

It is noteworthy that not dialect performance as a relatively objective and quantitative measure of a dialect usage but dialect proficiency, a subjective and qualitative evaluation of one's dialect mastery, predicted the robot's perceived trustworthiness. The ability to speak a dialect can be integral to one's self-image and contribute to the identification of oneself with a particular group or set of qualities. According to recent research, it is so-called self-essentialist reasoning, that is beliefs about the essence of one self, that underly the similarity-attraction effect (Chu et al., 2019). This reasoning focuses more on what one is and not on what one does; it is a personal characteristic that tends to be stable rather than situational or temporary in nature.

On a side note, participants who watched the video on a PC rated the standard German-speaking robot as more trustworthy and more competent, compared to participants working on a tablet or a mobile phone. This result indicates that, when examining human-robot interaction through video or audio stimuli, it is important to consider and control for the experimental device used. Possible reasons for the observed difference include different testing situations, such as doing the experiment at home on a PC or "on the go" on a mobile phone, which could have resulted in different distractions and response criteria, or differences in information processing on different screens (cf. Sweeney and Crestani, 2006; Wickens and Carswell, 2021). These factors could have potentially led to increased cognitive load on smaller screens and, consequently, to trust misplacement (Duffy and Smith, 2014).

4.2 Limitations of the study

It is worth noting that various intervening factors could have influenced our study. First, choosing a male voice might have affected the overall outcomes. Unlike in human-human interactions (Bonein and Serra, 2009; Slonim and Guillen, 2010), prior studies have shown that virtual assistants or robots with a male voice are generally viewed as more competent (Powers and Kiesler, 2006; Ernst and Herm-Stapelberg, 2020) and trustworthy (Behrens et al., 2018), although these ratings can be context-dependent (Andrist et al., 2015; Kraus et al., 2018). On the contrary, other recent research indicates that a female voice agent may be viewed as more likable, competent, or intelligent (Vega et al., 2019; Dong et al., 2020).

Second, due to social identification, people tend to rate voices of the same gender as more trustworthy (Crowelly et al., 2009) and perceive more psychological closeness to them (Eyssel et al., 2012). However, our research did not find evidence for this when using male voice stimuli exclusively. To resolve these contradictory results, more studies utilizing both male and female voices are necessary.

Third, dialects carry distinct connotations within German-speaking countries (cf. H. Bishop et al., 2005). For instance, the Berlin dialect is often associated with a lower socioeconomic class or working class (Stickel, 1997), whereas the Bavarian dialect is often viewed as more prestigious. It is even mandatory for politicians to speak the local dialect in Bavaria. In particular, the Bavarian dialect of Germany holds a significant and independent position within the conceptual framework of languages (Adler, 2019). A survey revealed that the Bavarian dialect is considered the second most appealing German dialect (29.6%), after the Northern German dialect (34.9%), while only about 7% found the Berlin dialect attractive (Gärtig et al., 2010; Adler and Plewnia, 2018). At the same time, a mere 5% of respondents found the Berlin dialect unappealing, whereas having no dialect at all was rated as unattractive by 32.6% of the participants (Gärtig et al., 2010). Thus, to obtain a more nuanced understanding, it would be beneficial to conduct a comparative study involving multiple dialects as well as add an assessment of subjective dialect connotations. Moreover, as dialects are a means of positive identification within a group and signify a sense of attachment to a particular region (Wiese, 2012), varying levels of identification may exist among different dialects. This can affect the degree of perceived similarity and subsequently influence assessments of trustworthiness and competence.

Fourth, our study employed a video featuring NAO, a compact and intelligent-looking social robot. It remains uncertain if its appearance aligns with all the connotations linked to the Berlin dialect. Humans may link voices with robots, and a mismatch in this connection could result in diverse outcomes in their interaction (McGinn and Torre, 2019).

Finally, we consider the limitations of our methodology for data collection and data analysis. With regard to data collection, it will be important to provide converging evidence for this internet-based study by conducting both laboratory-based and real-life research in future projects. With regard to data analysis, more advanced modeling techniques, like linear mixed modeling, can offer greater flexibility compared to stepwise regression and can usefully be employed to uncover additional effects in our data, including further variability driven by participant characteristics.

Also, the topic of communication can influence the assessment of a robot that speaks a particular dialect. Using standard German would likely be more suitable for discussing a painting, while a dialect such as the Berlin dialect could be more appropriate for conversations about everyday events or work-related topics (topic-based shifting) (Walker, 2019).

An overall point for future investigations is that certain scholars view trust as a construct that has multiple dimensions. For example, Law and Scheutz (2021) differentiate between performance-based trust and relation-based trust. Future research on trust should take into account these different aspects and explore their implications in various contexts. Finally, objective measures of trust, for example, following a robot’s advice or task delegation should be used to better operationalize the outcome (Law and Scheutz, 2021).

Overall, our study provides valuable insights into how language proficiency and other demographic factors influence human-robot interaction and robot perception. Our results can inform the development of more effective robots that are tailored to meet the needs and expectations of diverse user groups. Further research is needed to explore the role of gender, age, and dialect in human-robot interaction and perception, and to identify additional factors that may influence trust and competence evaluation.
Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found below: https://osf.io/pfqg6/.

Ethics statement

The study was conducted in accordance with the guidelines laid down in the Declaration of Helsinki and in compliance with the ethics policy of the University of Potsdam. No explicit approval was needed because the methods were standard. There were no known risks and participants gave their informed consent. The study and the procedure were already evaluated by professional psychologists to be consistent with the ethical standards of the German Research Foundation, including written informed consent and confidentiality of data as well as personal conduct.

Author contributions

KK and EH contributed to the conception and design of the study. EH conceived the stimuli, programmed the survey, and conducted the study. KK and EH performed the analysis. KK wrote the first draft of the manuscript. KK, MF, OB, and YZ wrote, discussed, and revised several drafts before approving the final version. All authors contributed to the article and approved the submitted version.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. Funded by the German Research Foundation (DFG) - Project number 491466077.

Acknowledgments

We would like to thank Tristan Kornher for creating and generously providing his video footage and Alexander Schank for providing his voice for the corresponding material.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher’s note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/frobt.2023.1241519/full#supplementary-material

References


