To trust or not to trust? Face and voice modulation of virtual avatars

Sebastian Siehl[1,2]*[†], Kornelius Kammler-Sücker[3,4†],
Stella Guldner[2], Yannick Janvier[4], Rabia Zohair[1] and Frauke Nees[1]

[1]Institute of Medical Psychology and Medical Sociology, University Medical Center Schleswig-Holstein, Kiel University, Kiel, Germany, [2]Department of Child and Adolescent Psychiatry and Psychotherapy, Central Institute of Mental Health, Medical Faculty Mannheim, Ruprecht-Karls-University Heidelberg, Mannheim, Germany, [3]Center for Innovative Psychiatric and Psychotherapeutic Research (CIPP), Central Institute of Mental Health, Medical Faculty Mannheim, Ruprecht-Karls-University Heidelberg, Mannheim, Germany, [4]Institute of Cognitive and Clinical Neuroscience, Central Institute of Mental Health, Medical Faculty Mannheim, Ruprecht-Karls-University Heidelberg, Mannheim, Germany

Introduction: This study explores the graduated perception of apparent social traits in virtual characters by experimental manipulation of perceived affiliation with the aim to validate an existing predictive model in animated whole-body avatars.

Methods: We created a set of 210 animated virtual characters, for which facial features were generated according to a predictive statistical model originally developed for 2D faces. In a first online study, participants (N = 34) rated mute video clips of the characters on the dimensions of trustworthiness, valence, and arousal. In a second study (N = 49), vocal expressions were added to the avatars, with voice recordings manipulated on the dimension of trustworthiness by their speakers.

Results: In study one, as predicted, we found a significant positive linear (p < 0.001) as well as quadratic (p < 0.001) trend in trustworthiness ratings. We found a significant negative correlation between mean trustworthiness and arousal (τ = −.37, p < 0.001), and a positive correlation with valence (τ = 0.88, p < 0.001). In study two, wefound a significant linear (p < 0.001), quadratic (p < 0.001), cubic (p < 0.001), quartic (p < 0.001) and quintic (p = 0.001) trend in trustworthiness ratings. Similarly, to study one, we found a significant negative correlation between mean trustworthiness and arousal (τ = −0.42, p < 0.001) and a positive correlation with valence (τ = 0.76, p < 0.001).

Discussion: We successfully showed that a multisensory graduation of apparent social traits, originally developed for 2D stimuli, can be applied to virtually animated characters, to create a battery of animated virtual humanoid male characters. These virtual avatars have a higher ecological validity in comparison to their 2D counterparts and allow for a targeted experimental manipulation of perceived trustworthiness. The stimuli could be used for social cognition research in neurotypical and psychiatric populations.

KEYWORDS

trustworthiness, virtual avatars, virtual reality, facial expressions, vocal expressions

1 Introduction

Humans are highly evolved social animals and constantly extract information about someone's trustworthiness from cues in their environment. The perception and evaluation of social information also becomes increasingly important in the "virtual world" given the ideas of a metaverse (Dionisio et al., 2013) or serious games (Charsky, 2010). Faces are valuable carriers and a primary source of information for seeming character traits (Kanwisher, 2000; Oosterhof and Todorov, 2008; Zebrowitz and Montepare, 2008). Facial features are quickly assessed by perceivers (Olivola and Todorov, 2010), first impressions of someone's personality are inferred in as quickly as 100 milliseconds (Zebrowitz, 2017b), and decisions are made on the basis of these evaluations with consequences reaching from the selection of a romantic partner to voting preferences (Olivola and Todorov, 2010; South Palomares and Young, 2018). Oosterhof and Todorov (2008) proposed a 2D social trait space model for face evaluation. This data-driven approach revealed two principal components, namely, apparent trustworthiness and dominance, explaining most of the variance in regard to social judgement of faces (Todorov et al., 2013). It is important to stress, in Todorov's own words, that these models merely describe "our natural propensity to form impressions" and should not lead into "the physiognomist' trap" of taking facial features "as a source of information about character" in the actual sense (Todorov, 2017, p. 27 and p. 268).

An equally important source of information consulted by social interaction partners is an individual's voice (Pisanski et al., 2016; Lavan et al., 2019). Similar emotional and social representations have been found for faces and voices (Kuhn et al., 2017). The information carried by flexible modulation of a speaker's voice is also used for the formation of first impressions, extracting social information about someone's trustworthiness and dominance (McAleer et al., 2014; Belin et al., 2017; Leongómez et al., 2021), which can be intentionally affected by speakers through targeted voice modulation (Hughes et al., 2014; Guldner et al., 2020).

Social traits were hitherto mostly studied using 2D images of isolated faces or vocal recordings. However, as humans, we mainly interact with multimodal 3D stimuli in the real as well as virtual world (e.g., computer gaming). Human-like virtual characters, so-called humanoid avatars, have previously been integrated and proved beneficial in multiple avenues of research and applied science (Kyrlitsias and Michael-Grigoriou, 2022) including educational support for children (Falloon, 2010), support for athletes (Proshin and Solodyannikov, 2018), studying social navigation in a virtual town (Tavares et al., 2015), rehabilitation training via a virtual coach (Birk and Mandryk, 2018; Tropea et al., 2019) or therapeutic interventions such as embodied self-compassion (Falconer et al., 2016).

Virtual reality (VR) technology in general, both in the strict sense of immersive VR (e.g., realized via VR helmets or head-mounted displays) and in the broad sense of any computer-generated virtual environment, is increasingly used in psychiatric and psychotherapeutic research on social perception and interaction (Falconer et al., 2016; Freeman et al., 2017; Maples-Keller et al., 2017). Virtual characters, as artificial interaction partners, provide this research with a form of (semi-)realistic social stimuli (Pan and Hamilton, 2018) that allow for standardized experimental setups with gradual modification of diverse features, such as realism and visual appearance. As such, they facilitate experimental control of social impressions evoked in human participants. The generalizability of behavior observed in the laboratory to natural behavior in the world, also called ecological validity (Schmuckler, 2001), has been shown to be increased in virtual environments and with 3D stimuli (Parsons, 2015; Kothgassner and Felnhofer, 2020). The immersiveness of VR provides a possible avenue for translating results from basic research and enhance complexity of stimulus material at the same time. Multisensory stimuli, e.g., simultaneous visual and auditory input, has long been discussed as an additional factor to increase ecological validity (De Gelder and Bertelson, 2003). A more complete evaluation of trustworthiness might therefore be achieved with virtual multisensory stimuli.

This study explores the gradual experimental manipulation of perceived personality traits in virtual characters with the aim of building an openly available databank of validated humanoid avatars. Target outcome was the perceived trustworthiness of 210 animated virtual characters, which were presented in short video clips on a 2D screen and rated by volunteers in online surveys. The first experiment presented here used modifications in facial features to elicit different levels of apparent trustworthiness, employing the statistical model by Todorov et al. (2013) and extending it to animated full-body characters. Our hypothesis was that the different levels of trustworthiness, as predicted by the model derived from static 2D images of virtual faces, would be confirmed in the visually more complex and dynamic stimuli as well (study 1). The second study combined these visual stimuli with auditory social information: the animated characters uttered sentences which carried little informational content but were shaped by emphatic social voice modulation of trustworthiness, with a set of auditory stimuli recorded from real-world speakers (Guldner et al., 2023 in preparation). Our hypothesis was that this multimodal manipulation would further enhance the differentiation of distinct levels of perceived trustworthiness for virtual characters, compared to unimodal visual stimuli.

2 Materials and methods

2.1 Participants

In study 1, 34 individuals ($N_{females} = 26$, $M_{age} = 23.0$, $SD_{age} = 4.3$) were recruited to take part in an online study, whereas in study 2, 49 individuals ($N_{females} = 38$, $M_{age} = 22.7$, $SD_{age} = 2.9$) participated. The following inclusion criteria were applied for both studies: neurotypical German speakers, age range of 18–35 years, normal or corrected-to-normal vision. Participants were recruited via advertisements on the website of the Central Institute of Mental Health in Mannheim as well as on social media. The participation in these online studies was completely voluntary and participants were informed prior of taking part that they could withdraw from the study at any point without any negative consequences and without having to give a reason. The studies were carried out in accordance with the Code of Ethics of the World Medical Association (World Medical Association, 2013) and were approved by the Ethical Review Board of the Medical Faculty Mannheim, Heidelberg University. All participants gave written (online) informed consent.

2.2 Stimuli and experimental procedure

2.2.1 Faces
2.2.1.1 Facegen

The stimuli were adapted from Todorov et al. (2013) face stimuli. We used 210 male Caucasian faces, created with a statistical model by Todorov et al. (2013) to modify facial expressions of computer-generated, emotionally neutral faces on the dimensions of trustworthiness, using the Facegen Modeller program ([Singular Inversions, Toronto, Canada] for more details, please review: Oosterhof and Todorov, 2008). The model was used to generate 30 new emotionally neutral faces, for each of which seven different variations were created by manipulating certain facial features on the dimension of trustworthiness. These variations ranged from "very trustworthy" to "very untrustworthy." They were achieved by moving one, two and three standard deviations (−3 SD [very untrustworthy] to +3 SD [very trustworthy]) away from the neutral face. This procedure was carried out for all 30 neutral faces, creating 30 face families, each having seven individual faces varying in trustworthiness, yielding a total of 210 (30 × 7) faces Figure 1.

2.2.1.2 Body modelling

We used ADOBE Fuse to model a template body which could be used with all face families. This was done to ensure that the avatars would only differ with respect to their faces. The modelled body represented a masculine body, dressed in a black turtleneck jacket and blue pants. Once it had been created, the body was exported as a Wavefront object (.obj) file Figure 1.

2.2.1.3 Rigging the body and adding animation

The process of rigging is used in 2D and 3D animation to build a model's skeletal structure to animate it. ADOBE Mixamo (https://www.mixamo.com), an online software, was used to add skeletons (rigging) and animations to the template body. We animated the latter (using a movement animation) so that it appeared as casually breathing. The animated template body was then downloaded in Autodesk Filmbox (FBX) file format Figure 1.

2.2.1.4 Joining head and body

Once we had both the Facegen heads and the animated body, we had to join them to create the avatars. This was accomplished using the software Autodesk 3ds Max (3ds Max 2019 | Autodesk, San Rafael, CA) using a customized script written in the programming language "Max Script," by author K.K.S. The script was used to optimize the process of creating the avatars by downloading all Facegen faces in OBJ format, inserting them on the animated template body in FBX format, connecting the head to the animated bones, and then re-exporting the whole object as a fully animated avatar in configuration (CFG) file format Figure 1.

2.2.1.5 Rendering the avatar

The avatars were then rendered, in a video format, using Vizard (WorldViz VR, Santa Barbara, CA, https://www.worldviz.com/vizard-virtual-reality-software), a virtual reality development software. Author K.K.S. programmed a customized script for this process which loaded each avatar in CFG file format, played the animation, recorded the screen, saved it as an Audio Video Interleave (AVI) file, and later converted it to an MP4 file. This

resulted in a total of 210 MP4 files, each having a duration of 7 seconds with a fully animated avatar. In other words, 30 animated avatar video families, each having seven individual avatars varying on the dimension of trustworthiness, were created Figure 1.

2.2.2 Vocal recordings

In study 2, we selected twelve avatars from study one, which showed linear trends across the dimension of trustworthiness. We then added voice recordings from a previous study [Guldner et al., 2023; procedure similar to Guldner et al., 2020]. The voice recordings were produced by non-trained speakers who modulated their voice to express likeability or hostility, or their neutral voice. This dimension is closely related to the trustworthiness dimension in the social voice space (Fiske et al., 2007) and was therefore used as a similar social content as the trustworthiness manipulation. The sentences were the following: "Many flowers bloom in July." (in German: "Viele Blumen blühen im Juli"), "Bears eat a lot of honey" (in German: "Bären essen viel Honig") and "There are many bridges in Paris." (in German: "Es gibt viele Brücken in Paris."). The voice recordings were further processed using the speech analysis software PRAAT (Boersma 2001) to filter out any breathing sounds, tongue clicking and throat clearing. The recordings were than rated by the speakers themselves for each relevant trait expression, i.e., likeability or hostility. Based on the self-ratings for each speaker, we selected voice recordings which varied in the intensity of likeability/hostility to match the avatar manipulations. For instance, recordings which received a speaker's maximal intensity in expressing a likeable voice were matched with maximal trustworthiness manipulation of the 3D Avatar, whereas the recording receiving maximal ratings for hostility was paired with the maximal untrustworthy avatar manipulation. Recordings from one speaker were assigned only to one avatar family, respectively. Thus, for this study, the neutrally expressed voices were assigned to the neutrally perceived faces and the different levels of expressed likeability/hostility in the voices were also assigned to the according level of perceived trustworthiness in the faces from study 1. As a result, we created congruent trials with matching levels of trustworthiness for voices and faces. Using 3ds Max and Vizard, lips and mouth movements were added to the avatars in synchrony to the voice recording, aiming at mimicking humanoid behavior.

2.2.3 Experimental design and procedure

In study 1, we had a 30 (avatar families) x 7 (avatar manipulations) repeated-measures design, whereas in study 2, it was reduced to a 12 (avatar families) x 7 (avatar manipulations) design. In both experiments, participants viewed all avatar families and manipulations sequentially in random order.

Both studies were conducted in SoSci survey (Leiner, 2019), a web-based platform for creating and distributing surveys. Participants could access the studies via a link to the SoSci survey platform (https://www.soscisurvey.de) from any computer system. For randomization of the stimuli, we relied partly on customized PHP and HTML scripts (author R.Z. and Y.J.). Participants were informed about the purpose of the study and the procedure before giving their consent for participation. In study 2, participants were presented with an exemplary voice recording saying "Paris" to test and adjust their volume. Participants were asked to name the city just mentioned in the recording on the next page and were instructed to adjust their volume if necessary, so they could hear the voice recording loud

and clear. They were then instructed to keep the volume on the same level for the rest of the experiment. During both experiments, participants were presented with several catch trials in between experimental trials to check whether they kept paying attention. The catch trials consisted of questions, e.g., about the appearance of the avatar in the last trial or the sentence they just heard.

Each video with the avatars was presented on a separate page. The video started automatically, and participants were instructed to look at the avatar for at least 4 seconds. After 4 seconds, the following four questions appeared below the video for rating: 1) "How trustworthy do you find the avatar? Please move the slider to the position that most accurately matches your attitude towards the virtual character." (in German: "Wie vertrauenswürdig finden Sie diese virtuelle Person? Bewegen Sie bitte den Schieberegler auf die Position, die am meisten ihrer Einstellung gegenüber der virtuellen Person entspricht."); 2) "How likeable do you find the virtual person as a whole?" (in German: "Wie sympathisch erscheinen Ihnen die virtuelle Person als Ganzes?"); 3) "How positive or negative do you find the avatar?" (in German: "Wie positiv oder negativ nehmen Sie die Person wahr?"); 4) "How exciting do you find the avatar?" (in German: "Wie aufregend nehmen Sie die Person wahr?"). Participants were not informed about the manipulation of the faces and voices in advance and were instructed to rely on their first impressions and to respond as quickly as possible.

Trustworthiness and likeability were rated on a visual analogue scale (VAS) ranging from "not at all trustworthy/likeable" to "extremely trustworthy/likeable." We chose a unipolar measure of trustworthiness with the VAS ratings ranging from "0" (not at all trustworthy) to "100" (extremely trustworthy). Arousal and valence were rated on a nine-point digital version of the "Self-Assessment mannikin" (SAM) scales (Bradley and Lang, 1994), consisting of nine full-body figures representing each dimension from "negative to positive" (valence) and "calm to excited" (arousal). The widely used SAM ratings were on a 9-point scale, where "0" was the negative extreme and "8" was the positive extreme. To limit the amount of experimental duration for participants, we limited the number of trials and did not separate the ratings of trustworthiness from arousal and valence for each avatar, thus they appeared on the same page.

2.3 Statistical analysis

All statistical analyses were performed in R-Statistics (Team, 2013). Data were assessed for outliers and normal distribution. Data was preprocessed with the dplyr package (Wickham et al., 2019), and figures were plotted with ggplot2 (Wickham, 2017). We used linear mixed effect models (lme) with a-priori defined contrasts (linear, quadratic). To control for variance caused by participant effects and avatar family in the response variable, we defined these factors/participant and avatar family as random effects (with only random intercepts, no random slopes), whereas an avatars' trustworthiness manipulation was defined as a fixed effect. All trends reported in the results section refer to fixed effects of contrasts for this variable. The contrasts were defined using the lsmeans package version 2.30-0 (Lenth, 2016), models were fitted with the lme4 package version 1.1-26 (Bates et al., 2015), and p-values

were estimated with the Satterthwaite method as implemented in the lmerTest package version 3.1-3 (Kuznetsova et al., 2017).

3 Results

3.1 Study 1

Figure 2 illustrates the trustworthiness ratings of participants in dependence of the avatars' manipulation. As predicted, we found a significant positive linear ($t_{(7071)} = 47.67$, $p < 0.001$) and quadratic ($t_{(7071)} = -4.76$, $p < 0.001$) trend in trustworthiness ratings. Furthermore, we found a significant positive linear ($t_{(7071)} = 52.96$, $p < 0.001$), as well as quadratic ($t_{(7071)} = -5.41$, $p < 0.001$), trend in likeability ratings (Figure 3; Table 1).

We found a significant negative correlation between mean trustworthiness ratings and mean arousal ratings across all avatar manipulations ($\tau = -.37$, $p < 0.001$), and a positive correlation between mean trustworthiness ratings and mean valence ratings ($\tau = .88$, $p < 0.001$; Figure 3 and Table 3).
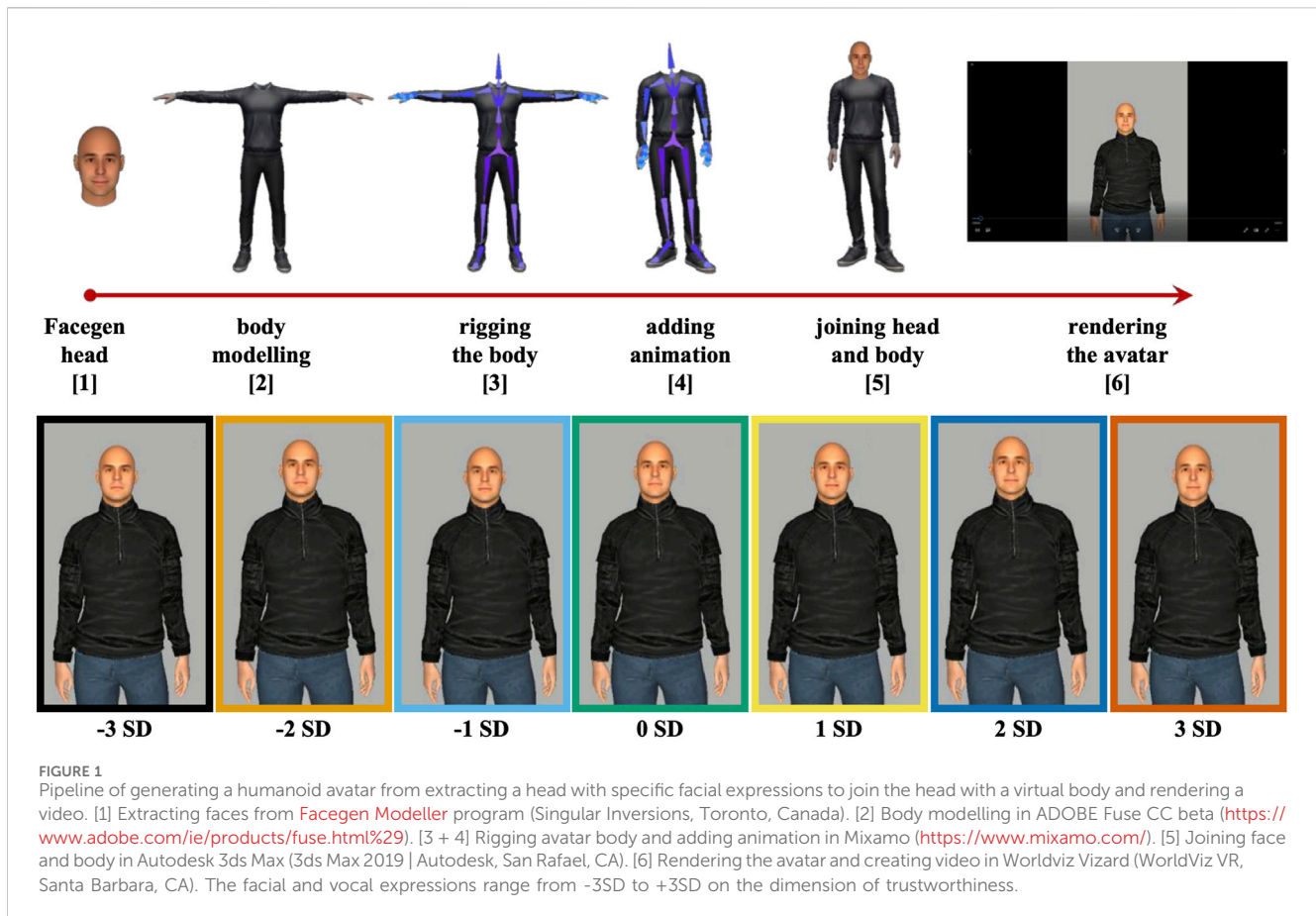
3.2 Study 2

In study two, we found a significant linear ($t_{(4465.96)} = 33.91$, $p < 0.001$), quadratic ($t_{(4465.96)} = -10.05$, $p < 0.001$), cubic ($t_{(4465.96)} = -5.90$, $p < 0.001$), quartic ($t_{(4465.96)} = 4.88$, $p < 0.001$) and quintic ($t_{(4465.96)} = 3.20$, $p = 0.001$) trend in trustworthiness ratings. For likeability, there were significant linear ($t_{(4465.96)} = 45.71$, $p < 0.001$), quadratic ($t_{(4465.96)} = -10.97$, $p < 0.001$), cubic ($t_{(4465.96)} = -8.18$, $p < 0.001$), quartic ($t_{(4465.96)} = 4.99$, $p < 0.001$) as well as sextic ($t_{(4465.96)} = -2.37$, $p = 0.02$) trends (Figure 2; Table 2).

Similarly to study one, we found a significant negative correlation between mean trustworthiness ratings and mean arousal ratings across all avatar manipulations ($\tau = -0.42$, $p < 0.001$), and a positive correlation between mean trustworthiness ratings and mean valence ratings ($\tau = 0.76$, $p < 0.001$; Figure 3 and Table 3).

4 Discussion

The present study investigated a unimodal and a multimodal method to manipulate the perceived trustworthiness of animated virtual characters in a highly controlled and gradual manner. The first experiment could establish that animated whole-body avatars are rated similarly to static faces by observers, and found that a gradual manipulation of facial features based on the model by Todorov et al. (2013) successfully predicted perceived trustworthiness with a linear relation. The second experiment could show that this effect of facial morphing is enhanced by targeted voice modulation of the characters' verbal utterances, with a significant role of nonlinear factors in the relationship between multimodal experimental manipulation and perceived trustworthiness. Besides these main findings, both experiments found high correlations between perceived social traits (trustworthiness) and emotional responses (valence and arousal).

FIGURE 1
Pipeline of generating a humanoid avatar from extracting a head with specific facial expressions to join the head with a virtual body and rendering a video. [1] Extracting faces from Facegen Modeller program (Singular Inversions, Toronto, Canada). [2] Body modelling in ADOBE Fuse CC beta (https://www.adobe.com/ie/products/fuse.html%29). [3 + 4] Rigging avatar body and adding animation in Mixamo (https://www.mixamo.com/). [5] Joining face and body in Autodesk 3ds Max (3ds Max 2019 | Autodesk, San Rafael, CA). [6] Rendering the avatar and creating video in Worldviz Vizard (WorldViz VR, Santa Barbara, CA). The facial and vocal expressions range from -3SD to +3SD on the dimension of trustworthiness.

The first experiment transferred a predictive model for perceived trustworthiness developed on static face stimuli to dynamic video stimuli of whole-body characters. This could predict perceived trustworthiness with a linear relation, confirming the relationship found in the original validation study by Todorov et al. (2013). These findings suggest that in this case, the transition towards higher realism did not evoke an "uncanny valley" effect of more realistic virtual agents evoking aversive reactions, such as revulsion and feelings of creepiness (Mori et al., 2012). Here, the lack in photorealism may have been advantageous with respect to the study goal of establishing gradual manipulations of trustworthiness *within* exemplars of the same type of stimuli. Participants did indeed differentiate diverse levels of apparent trustworthiness, adding another example for virtual characters as successful stimulators of social perceptions and reactions (de Borst and de Gelder, 2015). A transfer from screen-based applications towards immersive virtual reality setups may further increase the perceived realism of the virtual characters explored in this study, possibly further enhancing their perceptual and behavioral effects in real-world interaction partners. Offering gradual and quantifiable experimental control of seeming trustworthiness, these visual stimuli could be applied fruitfully in different research areas. Among these are experimental psychology research on social decision making, which could use virtual characters as interaction partners in trust games (Pan and Steed, 2017; Krueger and Meyer-Lindenberg, 2019), as well as investigations of the neural correlates of navigation through social

spaces (Tavares et al., 2015; Schafer and Schiller, 2018; Zhang et al., 2022), which could be warped experimentally by regulating the level of seeming trustworthiness of virtual partners. In addition, research on ingroup-outgroup distinctions and the formation and dissolution of stereotypes could let participants embody virtual characters of different seeming trustworthiness, by letting them control the characters themselves and possibly even see themselves as these characters in a virtual mirror in immersive VR. These are the presuppositions of the *Proteus* effect (Yee and Bailenson, 2007), which describes a shift in behavioral and perceptual attitudes towards virtual characters (and humans similar to these characters) after embodying them. In our case, the *Proteus* effect after impersonating avatars with seemingly untrustworthy faces might attenuate participants' adverse reactions to these facial features. The allocation of trust towards others is often changed in characteristic ways in psychological and psychiatric conditions, as, for example, in borderline personality disorder (Franzen et al., 2011; Fertuck et al., 2013), and in people who went through several adverse childhood experiences (Hepp et al., 2021; Neil et al., 2021); here, virtual characters that are standardized in their average elicitation of perceived trustworthiness could provide clinical research with a valuable tool for exploring the distinct patterns of trust allocation in different disorders (Schilbach, 2016; Redcay and Schilbach, 2019).

The second experiment added auditory stimuli which were characterized by deliberate voice modulation of their original speakers, which had aimed at specific levels of (un-)
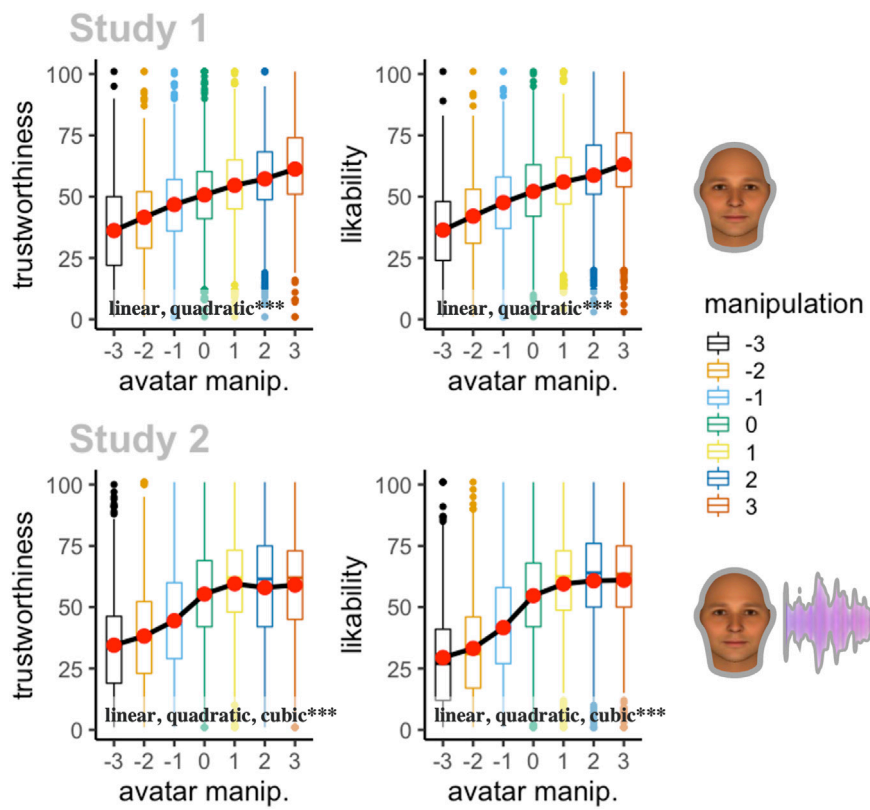
FIGURE 2
Boxplots (median, two hinges and outliers) depicting ratings of trustworthiness and liability as a function of an avatar's trustworthiness manipulation for study 1 and study 2. Furthermore, the linear trend of the mean of each rating is shown. Significant polynomial contrasts are depicted for each model.
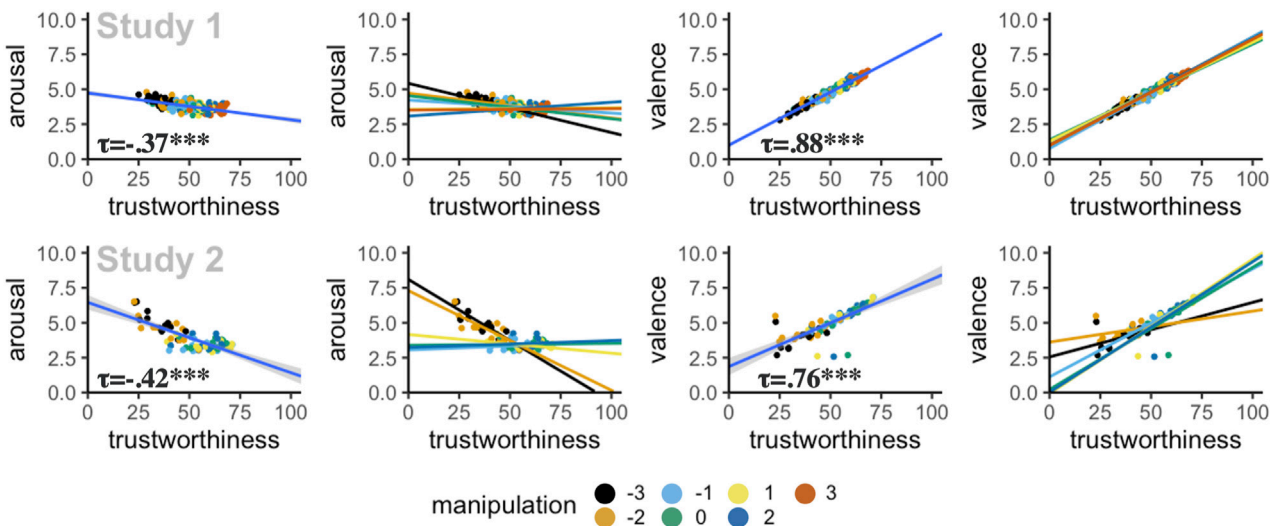

FIGURE 3
Scatterplots of Kendall's rank correlations between mean arousal/valence and trustworthiness ratings for study 1 and study 2. Each correlation is depicted as mean over all avatar manipulation and separately for each manipulation.

trustworthiness. The auditory stimuli had been chosen to match the level of trustworthiness of the avatars, according to intentionally modulated voice recordings of untrained speakers and their own self-ratings (Guldner et al., 2023 in preparation). Twelve avatar families had been selected, based on a maximally differentiated rating of trustworthiness in the first study. The combination of these

TABLE 1 Linear mixed effects model of study 1.

Study 1				
Trustworthiness				
parameter	b^a	SE^b	t	p
intercept	49.76	1.67	29.48	<0.001
linear	21.62	0.45	47.67	<0.001
quadratic	−2.16	0.45	−4.76	<0.001
cubic	0.63	0.45	1.39	0.17
x^4	0.51	0.45	1.13	0.26
x^5	0.16	0.45	0.36	0.72
x^6	0.35	0.45	0.78	0.44
Likeability				
parameter	b	SE	t	p
intercept	50.84	1.73	29.41	<0.001
linear	23.06	0.44	52.96	<0.001
quadratic	−2.36	0.44	−5.41	<0.001
cubic	0.69	0.44	1.58	0.11
x^4	0.75	0.44	1.73	0.08
x^5	0.22	0.44	0.50	0.62
x^6	0.21	0.44	0.48	0.63

Results of linear mixed effects model of study 1 using polynomial contrasts.
[a]beta estimate.
[b]standard error.

TABLE 2 Linear mixed effects model of study 2.

Study 2				
Trustworthines				
parameter	b	SE	t	p
intercept	49.87	2.23	22.39	<0.001
linear	24.27	0.72	33.91	<0.001
quadratic	−7.19	0.72	−10.05	<0.001
cubic	−4.22	0.72	−5.90	<0.001
x^4	3.50	0.72	4.88	<0.001
x^5	2.29	0.72	3.20	0.001
x^6	−0.99	0.72	−1.38	0.17
Likeability				
parameter	b	SE	t	p
intercept	48.59	2.22	21.94	<0.001
linear	31.71	0.69	45.71	<0.001
quadratic	−7.61	0.69	−10.97	<0.001
cubic	−5.67	0.69	−8.18	<0.001
x^4	3.46	0.69	4.99	<0.001
x^5	1.17	0.69	1.69	0.09
x^6	−1.64	0.69	−2.37	0.02

Results of linear mixed effects model of study 2 using polynomial contrasts.
[a]beta estimate.
[b]standard error.

two stimuli sets generated multimodal stimuli for which the prediction of apparent trustworthiness was also successful. The link between different manipulation levels and perceived trustworthiness, however, revealed a more complex pattern with significant nonlinear relations: higher polynomial orders of manipulation level were significant predictors of trustworthiness ratings. The global relation between manipulation levels and ratings resembles a stimulus-response curve with a saturation effect for positive trustworthiness levels. This could possibly be explained by a boundary effect of overdetermination by congruent multimodal social cues. A nonlinear stimulus-response curve with broad plateaus of (un-)trustworthiness could also reflect an adaptive function of social estimations from first impressions. These play a pivotal role in social decision-making, which regularly requires nonlinear behavioral outcomes in form of decisions between discrete alternatives: for example, if a social situation comes down to requiring a "yes-or-no" decision, a trustworthiness estimate is necessarily transformed into a binary outcome variable at some point during the decision-making process. Hence, the nonlinear factors in the relation between manipulation level and trustworthiness ratings potentially indicate that these multimodal stimuli further approached the behavioral significance and thereby perceptual salience of real-world social interaction partners. Note, however, that mean trustworthiness ratings in our studies were limited to "saturate" below a trustworthiness level of roughly 75%

TABLE 3 Correlations.

Study 1			
	τ^a	z	p
Arousal ρ Trustworthiness	−0.37	−7.86	<0.001
Valence ρ Trustworthiness	0.88	18.79	<0.001
Study 2			
	τ	z	p
Arousal ρ Trustworthiness	−0.42	−5.14	<0.001
Valence ρ Trustworthiness	0.76	9.43	<0.001

Results of Kendall's rank correlations.
[a]Kendall's tau as correlation coefficient.

for all avatar families, indicating an upper boundary to effects of our multimodal manipulations. This suggests that these "carry only so far" in facilitating trustworthiness, given the generic common characteristics of all our stimuli, as, e.g., all representing bald males clothed in dark colors without ample emotional expression.

Another indicator for an increased complexity of interactions between multimodal cues is the variety of shapes in stimulus-response curves for the individual avatar families. For multimodal visual-auditory stimuli, these take many forms from

linear curves (resembling those for unimodal character stimuli) to almost stepwise sigmoidal jump functions, with the location of the transition between plateaus varying between avatar families (Supplementary Figure S2). This indicates that multimodality and increased realism can enhance behavioral responses to virtual characters but also render the interpretation of effects more challenging. Adding more layers of social information in future studies could further extend this research, such as by targeted variations in characters' gestures or clothing (Oh et al., 2019), more meaningful content of their utterances, and by narrative contextualization of characters with short fictional biographic episodes. These manipulations could further enhance the usability of virtual characters in psychological, psychiatric, and psychotherapeutic research.

4.1 Limitations

There are several limitations of this study. Firstly, our stimuli consisted of only white male virtual characters of roughly medium age, a design decision mainly due to similarly limited scope of the predictive model used to generate the stimuli. This certainly limits the generalizability of our findings, and future studies should extend the diversity of similar stimuli samples. Secondly, the voice recordings employed in the multimodal stimuli set were extracted from a study that required speakers to utter sentences with insignificant meaning. While this has certain advantages for evaluating social voice modulation independent of semantic content, the latter is usually an important cue for apparent trustworthiness in real-world social encounters. To increase ecological validity of future studies with virtual characters, the content of utterances should match the virtual social contexts of their expression. A third limitation of our study is the lack of behaviorally relevant interaction measures of trust based on, e.g., decision making in trust games. Future studies, for example, in the field of decision making, could make use of fully immersive virtual encounter situations that call for such decisions of trust on behalf of the participants. Fourthly, since trustworthiness, valence and arousal were rated on the same page by participants, the ratings could be influenced by each other. Lastly, the vocal recordings were intentionally manipulated on the likeability/hostility dimension of the social space, but not explicitly on trustworthiness. However, these dimensions are closely related on a perceptual level (McAleer et al., 2014). In fact, this was reflected in the highly similar rating profiles across trustworthiness and likeability ratings, in study 1 ($r_\tau = 0.89$, p=<0.001) and study 2 ($r_\tau = 0.86$, p= <0.001), supporting a common underlying dimension reflecting approach/avoidance behaviors to others (e.g., Fiske et al., 2007; Oosterhof and Todorov, 2008; Zebrowitz, 2017a).

4.2 Conclusion

In this cross-sectional study, we successfully showed that a multisensory graduation of social traits, originally developed for 2D stimuli, can be applied to virtually animated characters, to create a battery of animated virtual humanoid male characters. These allow for a targeted experimental manipulation of perceived trustworthiness of avatars with high ecological validity, which can be used in future desktop-based or fully

immersive virtual reality experiments. The predictive model was applied to visual facial features in the first study, which revealed significant linear trends in perceived trustworthiness. In the second study, which combined these with auditory vocal features, nonlinear trends in perceived trustworthiness gained more importance, possibly reflecting higher complexity of multimodal manipulations. These virtual avatars could provide a starting point for a larger data bank of validated multimodal virtual stimuli. This could be used for social cognition research in neurotypical and psychiatric populations.

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/jucza/?view_only=4ea3b5f449f940e1b571aeac0336f921.

Ethics statement

The studies involving humans were approved by Ethical Review Board of the Medical Faculty Mannheim, Heidelberg University. The studies were conducted in accordance with the local legislation and institutional requirements. The participants provided their written informed consent to participate in this study.

Author contributions

SS: Conceptualization, Formal Analysis, Investigation, Methodology, Project administration, Resources, Supervision, Visualization, Writing–original draft, Writing–review and editing. KK-S: Conceptualization, Methodology, Project administration, Resources, Software, Supervision, Visualization, Writing–original draft, Writing–review and editing. SG: Conceptualization, Investigation, Methodology, Resources, Supervision, Writing–original draft, Writing–review and editing. YJ: Investigation, Methodology, Writing–review and editing. RZ: Conceptualization, Data curation, Formal Analysis, Investigation, Visualization, Writing–original draft, Writing–review and editing. FN: Conceptualization, Funding acquisition, Resources, Supervision, Writing–review and editing.

Funding

Acknowledgments

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/frvir.2024.1301322/full#supplementary-material

References

Bates, D., Mächler, M., Bolker, B. M., and Walker, S. C. (2015). Fitting linear mixed-effects models using lme4. *J. Stat. Softw.* 67. doi:10.18637/jss.v067.i01

Belin, P., Boehme, B., and McAleer, P. (2017). The sound of trustworthiness: acoustic-based modulation of perceived voice personality. *PLoS One* 12, e0185651. doi:10.1371/journal.pone.0185651

Birk, M. V., and Mandryk, R. L. (2018). "Combating attrition in digital self-improvement programs using avatar customization," in Conf. Hum. Factors Comput. Syst. - Proc., Montreal QC Canada, April, 2018. doi:10.1145/3173574.3174234

Boersma, P. (2001). Praat, a system for doing phonetics by computer. *Glot International* 5, 341–345.

Bradley, M. M., and Lang, P. J. (1994). Measuring emotion: the self-assessment manikin and the semantic differential. *J. Behav. Ther. Exp. Psychiatry* 25, 49–59. doi:10.1016/0005-7916(94)90063-9

Charsky, D. (2010). From edutainment to serious games: a change in the use of game characteristics. *Games Cult.* 5, 177–198. doi:10.1177/1555412009354727

de Borst, A. W., and de Gelder, B. (2015). Is it the real deal? Perception of virtual characters versus humans: an affective cognitive neuroscience perspective. *Front. Psychol.* 6, 576. doi:10.3389/fpsyg.2015.00576

De Gelder, B., and Bertelson, P. (2003). Multisensory integration, perception and ecological validity. *Trends Cogn. Sci.* 7, 460–467. doi:10.1016/j.tics.2003.08.014

Dionisio, J. D. N., Burns, W. G., and Gilbert, R. (2013). 3D virtual worlds and the metaverse: current status and future possibilities. *ACM Comput. Surv.* 45, 1–38. doi:10.1145/2480741.2480751

Falconer, C. J., Rovira, A., King, J. A., Gilbert, P., Antley, A., Fearon, P., et al. (2016). Embodying self-compassion within virtual reality and its effects on patients with depression. *BJPsych Open* 2, 74–80. doi:10.1192/bjpo.bp.115.002147

Falloon, G. (2010). Using avatars and virtual environments in learning: what do they have to offer? *Br. J. Educ. Technol.* 41, 108–122. doi:10.1111/j.1467-8535.2009.00991.x

Fertuck, E. A., Grinband, J., and Stanley, B. (2013). Facial trust appraisal negatively biased in borderline personality disorder. *Psychiatry Res.* 207, 195–202. doi:10.1016/j.psychres.2013.01.004

Fiske, S. T., Cuddy, A. J. C., and Glick, P. (2007). Universal dimensions of social cognition: warmth and competence. *Trends Cogn. Sci.* 11, 77–83. doi:10.1016/j.tics.2006.11.005

Franzen, N., Hagenhoff, M., Baer, N., Schmidt, A., Mier, D., Sammer, G., et al. (2011). Superior "theory of mind" in borderline personality disorder: an analysis of interaction behavior in a virtual trust game. *Psychiatry Res.* 187, 224–233. doi:10.1016/j.psychres.2010.11.012

Freeman, D., Reeve, S., Robinson, A., Ehlers, A., Clark, D., Spanlang, B., et al. (2017). Virtual reality in the assessment, understanding, and treatment of mental health disorders. *Psychol. Med.* 47, 2393–2400. doi:10.1017/S003329171700040X

Guldner, S., Lavan, N., Lally, C., Wittmann, L., Nees, F., Flor, H., et al. (2023). Human talkers change their voices to elicit specific trait percepts. *Psychonomic Bulletin and Review.* doi:10.3758/s13423-023-02333-y

Guldner, S., Nees, F., and McGettigan, C. (2020). Vocomotor and social brain networks work together to express social traits in voices. *Cereb. Cortex* 30, 6004–6020. doi:10.1093/cercor/bhaa175

Hepp, J., Schmitz, S. E., Urbild, J., Zauner, K., and Niedtfeld, I. (2021). Childhood maltreatment is associated with distrust and negatively biased emotion processing. *Borderline Personal. Disord. Emot. Dysregulation* 8, 5–14. doi:10.1186/s40479-020-00143-5

Hughes, S. M., Mogilski, J. K., and Harrison, M. A. (2014). The perception and parameters of intentional voice manipulation. *J. Nonverbal Behav.* 38, 107–127. doi:10.1007/s10919-013-0163-z

Kanwisher, N. (2000). Domain specificity in face perception. *Nat. Neurosci.* 3, 759–763. doi:10.1038/77664

Kothgassner, O. D., and Felnhofer, A. (2020). Does virtual reality help to cut the Gordian knot between ecological validity and experimental control? *Ann. Int. Commun. Assoc.* 44, 210–218. doi:10.1080/23808985.2020.1792790

Krueger, F., and Meyer-Lindenberg, A. (2019). Toward a model of interpersonal trust drawn from neuroscience, psychology, and economics. *Trends Neurosci.* 42, 92–101. doi:10.1016/j.tins.2018.10.004

Kuhn, L. K., Wydell, T., Lavan, N., McGettigan, C., and Garrido, L. (2017). Similar representations of emotions across faces and voices. *Emotion* 17, 912–937. doi:10.1037/emo0000282

Kuznetsova, A., Brockhoff, P. B., and Christensen, R. H. B. (2017). lmerTest package: tests in linear mixed effects models. *J. Stat. Softw.* 82, 1–26. doi:10.18637/JSS.V082.I13

Kyrlitsias, C., and Michael-Grigoriou, D. (2022). Social interaction with agents and avatars in immersive virtual environments: a survey. *Front. Virtual Real* 2, 1–13. doi:10.3389/frvir.2021.786665

Lavan, N., Burton, A. M., Scott, S. K., and McGettigan, C. (2019). Flexible voices: identity perception from variable vocal signals. *Psychon. Bull. Rev.* 26, 90–102. doi:10.3758/s13423-018-1497-7

Lenth, R. V. (2016). Least-squares means: the R package lsmeans. *J. Stat. Softw.* 69. doi:10.18637/jss.v069.i01

Leongómez, J. D., Pisanski, K., Reby, D., Sauter, D., Lavan, N., Perlman, M., et al. (2021). Voice modulation: from origin and mechanism to social impact. *Philos. Trans. R. Soc. B Biol. Sci.* 376, 20200386. doi:10.1098/rstb.2020.0386

Maples-Keller, J. L., Bunnell, B. E., Kim, S. J., and Rothbaum, B. O. (2017). The use of virtual reality technology in the treatment of anxiety and other psychiatric disorders. *Harv. Rev. Psychiatry* 25, 103–113. doi:10.1097/HRP.0000000000000138

McAleer, P., Todorov, A., and Belin, P. (2014). How do you say "hello"? Personality impressions from brief novel voices. *PLoS One* 9, e90779–9. doi:10.1371/journal.pone.0090779

Mori, M., MacDorman, K. F., and Kageki, N. (2012). The uncanny valley [from the field]. *IEEE Robot. Autom. Mag.* 19, 98–100. doi:10.1109/MRA.2012.2192811

Neil, L., Viding, E., Armbruster-Genc, D., Lisi, M., Mareshal, I., Rankin, G., et al. (2021). Trust and childhood maltreatment: evidence of bias in appraisal of unfamiliar faces. *J. Child. Psychol. Psychiatry Allied Discip.* 63, 655–662. doi:10.1111/jcpp.13503

Oh, D. W., Shafir, E., and Todorov, A. (2019). Economic status cues from clothes affect perceived competence from faces. *Nat. Hum. Behav.* 4, 287–293. doi:10.1038/s41562-019-0782-4

Olivola, C. Y., and Todorov, A. (2010). Elected in 100 milliseconds: appearance-based trait inferences and voting. *J. Nonverbal Behav.* 34, 83–110. doi:10.1007/s10919-009-0082-1

Oosterhof, N. N., and Todorov, A. (2008). The functional basis of face evaluation. *Proc. Natl. Acad. Sci. U. S. A.* 105, 11087–11092. doi:10.1073/pnas.0805664105

Pan, X., and Hamilton, A. F. de C. (2018). Why and how to use virtual reality to study human social interaction: the challenges of exploring a new research landscape. *Br. J. Psychol.* 109, 395–417. doi:10.1111/bjop.12290

Pan, Y., and Steed, A. (2017). The impact of self-avatars on trust and collaboration in shared virtual environments. *PLoS One* 12, e0189078. doi:10.1371/journal.pone.0189078

Parsons, T. D. (2015). Virtual reality for enhanced ecological validity and experimental control in the clinical, affective and social neurosciences. *Front. Hum. Neurosci.* 9, 660–719. doi:10.3389/fnhum.2015.00660

Pisanski, K., Cartei, V., McGettigan, C., Raine, J., and Reby, D. (2016). Voice modulation: a window into the origins of human vocal control? *Trends Cogn. Sci.* 20, 304–318. doi:10.1016/j.tics.2016.01.002

Proshin, A. P., and Solodyannikov, Y. V. (2018). Physiological avatar technology with optimal planning of the training process in cyclic sports. *Autom. Remote Control* 79, 870–883. doi:10.1134/S0005117918050089

Redcay, E., and Schilbach, L. (2019). Using second-person neuroscience to elucidate the mechanisms of social interaction. *Nat. Rev. Neurosci.* 20, 495–505. doi:10.1038/s41583-019-0179-4

Schafer, M., and Schiller, D. (2018). Navigating social space. *Neuron* 100, 476–489. doi:10.1016/j.neuron.2018.10.006

Schilbach, L. (2016). Towards a second-person neuropsychiatry. *Philos. Trans. R. Soc. B Biol. Sci.* 371, 20150081. doi:10.1098/rstb.2015.0081

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

South Palomares, J. K., and Young, A. W. (2018). Facial first impressions of partner preference traits: trustworthiness, status, and attractiveness. *Soc. Psychol. Personal. Sci.* 9, 990–1000. doi:10.1177/1948550617732388

Tavares, R. M., Mendelsohn, A., Grossman, Y., Williams, C. H., Shapiro, M., Trope, Y., et al. (2015). A map for social navigation in the human brain. *Neuron* 87, 231–243. doi:10.1016/j.neuron.2015.06.011

Team, R. (2013). R: a language and environment for statistical computing. Available at: https://ftp.uvigo.es/CRAN/web/packages/dplR/vignettes/intro-dplR.pdf%20 (Accessed November 2, 2019).

Todorov, A. (2017). *Face value: the irresistible influence of first impressions*. Princeton, New Jersey, United States: Princeton University Press.

Todorov, A., Dotsch, R., Porter, J. M., Oosterhof, N. N., and Falvello, V. B. (2013). Validation of data-driven computational models of social perception of faces. *Emotion* 13, 724–738. doi:10.1037/a0032335

Tropea, P., Schlieter, H., Sterpi, I., Judica, E., Gand, K., Caprino, M., et al. (2019). Rehabilitation, the great absentee of virtual coaching in medical care: scoping review. *J. Med. Internet Res.* 21, e12805. doi:10.2196/12805

Wickham, H. (2017). ggplot2 - elegant graphics for data analysis (2nd edition). *J. Stat. Softw.* 77, 3–5. doi:10.18637/jss.v077.b02

Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L., François, R., et al. (2019). Welcome to the tidyverse. *J. Open Source Softw.* 4, 1686. doi:10.21105/joss.01686

World Medical Association, (2013). World Medical Association Declaration of Helsinki: ethical principles for medical research involving human subjects. *JAMA* 310, 2191–2194. doi:10.1001/jama.2013.281053

Yee, N., and Bailenson, J. (2007). The proteus effect: the effect of transformed self-representation on behavior. *Hum. Commun. Res.* 33, 271–290. doi:10.1111/j.1468-2958.2007.00299.x

Zebrowitz, L. A. (2017a). First impressions from faces. *Curr. Dir. Psychol. Sci.* 26, 237–242. doi:10.1177/0963721416683996

Zebrowitz, L. A. (2017b). First impressions from faces. *Curr. Dir. Psychol. Sci.* 26, 237–242. doi:10.1177/0963721416683996

Zebrowitz, L. A., and Montepare, J. M. (2008). Social psychological face perception: why appearance matters. *Soc. Personal. Psychol. Compass* 2, 1497–1517. doi:10.1111/j.1751-9004.2008.00109.x

Zhang, L., Chen, P., Schafer, M., Zheng, S., Chen, L., Wang, S., et al. (2022). A specific brain network for a social map in the human brain. *Sci. Rep.* 12, 1773–1816. doi:10.1038/s41598-022-05601-4