Reducing Racial Bias by Interacting with Virtual Agents: An Intervention in Virtual Reality

David Obremski
University of Würzburg
Würzburg, Germany
david.obremski@uni-wuerzburg.de

Ohenewa Bediako Akuffo
University of Würzburg
Würzburg, Germany
ohenewa.bediako_akuffo@stud-mail.uni-wuerzburg.de

Leonie Lücke
University of Würzburg
Würzburg, Germany
leonie.luecke@stud-mail.uni-wuerzburg.de

Miriam Semineth
University of Würzburg
Würzburg, Germany
miriam.semineth@stud-mail.uni-wuerzburg.de

Sarah Tomiczek
University of Würzburg
Würzburg, Germany
sarah.tomiczek@stud-mail.uni-wuerzburg.de

Hanna-Finja Weichert
University of Würzburg
Würzburg, Germany
hanna-finja.weichert@stud-mail.uni-wuerzburg.de

Birgit Lugrin
University of Würzburg
Würzburg, Germany
birgit.lugrin@uni-wuerzburg.de

Figure 1: The experimental setup. A) shows the experimenter controlling the Wizard-of-Oz interface. B) shows the two IVAs from the participant's point of view in Virtual Reality. C) shows a potential participant during the experiment.

ABSTRACT

Racial bias, implicit or explicit, is still a widespread phenomenon in our modern society, negatively affecting the way we interact with foreigners. These biases can also lead to the general tendency to avoid encounters with foreigners, which has a critical impact on society as a whole. This paper presents an approach to reduce implicit and explicit racial bias using two Intelligent Virtual Agents (IVAs) in Virtual Reality (VR). Based on previous research from social psychology and the field of enculturated IVAs, a sympathetic East African-German mixed-cultural IVA, and an antipathetic German mono-cultural IVA were implemented to interact with the participants in a virtual pub quiz. Pre- and post-intervention measures of participants' implicit and explicit bias showed a significant decrease in both scores. This work demonstrates the capability of enculturated IVAs to reduce real-world racial biases and sets the base for cultural interventions with IVAs.

KEYWORDS

Racial Bias; Intelligent Virtual Agents; Culture; Mixed-Cultural

1 INTRODUCTION

While meeting people of different cultural backgrounds was limited to holiday and business trips in the past, it has long since become a part of our everyday lives. Globalisation and migration contribute to an increasing cultural diversity worldwide, which leads to both

manifold opportunities as well as new and demanding challenges. One challenge people face when migrating to a new country is the adaptation to the new host culture. This adaptation usually manifests itself by partially taking on customs and beliefs of the host culture while also incorporating traits of the original culture. This leads to the development of a mixed-cultural background that is often easily recognised by members of the host culture, due to its distinct markers in the respective person's appearance, non-verbal behaviour and speech.

Since some of these cultural markers are difficult to conceal (e.g., non-native accents when learning a new language), mixed-cultural individuals have to deal with being perceived as such. Recent events in Germany have shown that not all mixed-cultural individuals are perceived and treated alike. While the public support for Ukrainian refugees after the start of the war in 2022 is very high in Germany, people seem to show less empathy towards refugees from more distant countries, as during the refugee crisis in 2015. This is in line with research on the behaviour towards cultural outgroups versus cultural ingroups, which reports a more positive perception of the latter [6].

With a total of 1.2 million refugees from mostly Arabic and East African countries that migrated to Germany between 2014 and 2021 [7, 46, 54], it is very relevant to mitigate prejudices and biases towards people of these less familiar cultural groups in Germany. While some might explicitly express their racial bias towards these mixed-cultural individuals, others might have an implicit racial bias without being aware of it. Both forms of racial bias are known to negatively affect one's behaviour towards foreigners [13, 14] which results in subsequent negative effects on society as a whole. Therefore, it is important to address this issue in education and interventions to raise cultural awareness.

However, cultural interventions with real human interlocutors are costly, not scalable and unsteady, as they rely on the consistency of the confederates. Intelligent Virtual Agents (IVAs) in contrast, represent a promising form of interaction partners, since they are cost-effective to design, and provide training scenarios without exposing the participant. Additionally, solutions with IVAs that raise cultural awareness can be distributed on a large scale.

Building on previous research from the social sciences and the field of enculturated IVAs, this paper describes the implementation of a prototype of a VR-based intervention with one sympathetic East African-German mixed-cultural IVA and one antipathetic German mono-cultural IVA, and investigates whether participants' implicit and explicit bias can be reduced.

2 THEORETICAL BACKGROUND OF SOCIAL PSYCHOLOGY

In their daily life, people tend to differentiate between members of their ingroup and outgroup members. Ingroup members are usually rated high on the two dimensions warmth and competence while outgroup members are only rated high on one of the two dimensions [59]. In general, ingroup members, whether the shared identity is based on religion [53], politics [45], or even on the random assignment to a group [55], are perceived more positively than outgroup members, which is known as *ingroup favoritism* [15, 58]. As this also holds true for people belonging to cultural outgroups

[6, 61], foreigners often have to deal with prejudice and implicit or explicit racial bias when moving to a new country. Since these attitudes do not only have negative effects on the individuals in question [13, 14], but also on the respective society's social cohesion, research in the social sciences has investigated how these biases can be mitigated (see [14] for an overview). In modern society, explicitly formulating negative statements about Blacks or people of colour is generally avoided because of social desirability and most people's self perception of being egalitarian and non-prejudiced [12].

While the mitigation of these rarely formulated explicit racial biases is usually tackled by presenting counter-stereotypical information about the respective group [12, 50], the mitigation of implicit racial biases is more complex. Since most people are unaware of potentially having an implicit racial bias, the mere assessment of it is challenging as self-reports or questionnaires can't be used. The most established way to measure implicit biases of any kind is to use the Implicit Association Test (IAT), which was developed by Greenwald et al. [16]. It measures the differential association of two target concepts with an attribute by assessing the participants reaction times. The concept behind the test is that the more associated the two presented stimuli are, the shorter the participants' reaction time is. Using the IAT as a measure, numerous attempts have been made to reduce implicit racial bias, as outlined in the following subsections. While the attempts to reduce implicit racial bias in the area of social psychology are manifold (e.g., through imagining contact with the cultural outgroup, appeals to egalitarian values, or inducing of emotion [14]), the following section focuses on the exposure to counter-stereotypical exemplars and the identification with the cultural outgroup since these two approaches could especially profit from the interactive nature of IVA-based interventions.

2.1 Exposure to Counter-Stereotypical Exemplars

One promising approach to reduce implicit racial bias seems to be the exposure to counter-stereotypical exemplars of the respective group. While the sole exposure of participants to positive exemplars of a different cultural background has been shown to effectively reduce implicit racial bias in some interventions [9, 34], it appeared to not be sufficient in others [30]. Adding negative examples of the participants' cultural ingroup however, seems to result in a more reliable reduction of implicit racial bias. Joy-Gaba and Nosek showed in an experiment that by exposing participants to admired Black and negative White exemplars the participants' implicit preference for Whites did significantly decrease [26]. A study reported by Lai et al. successfully reduced implicit racial bias by letting the participant play a dodge ball game in which all of their teammates were Black and positive and all of their opponents were White and negative [30].

2.2 Identification With the Cultural Outgroup

Another approach that has proven to effectively reduce implicit racial bias is the increased identification with cultural outgroups. Brannon and Walton conducted a study in which Americans interacted with Latino-American confederates that pretended to share

previously assessed interests with the participants to evoke the feeling of social connectedness despite being a member of a cultural outgroup [5]. The results reveal that the intervention with the confederates triggering social connectedness led to a significant decrease of the participants' implicit racial bias towards Latino-Americans. Another study that aimed at increasing the identification with a cultural outgroup was conducted by Hall et al. [19]. In the experiment participants were asked to list overlapping characteristics they would assign to both White and Black people. The authors showed that this task led to blurred intergroup boundaries between the participants and the cultural outgroup, which resulted in a significant decrease of implicit racial bias towards Blacks. Woodcock and Monteith were able to show that even by generating a fictional new ingroup by putting White participants in teams with Black individuals before completing a computer-mediated task, the participants' implicit racial bias towards Blacks decreased significantly [62].

Conclusion from the Theoretical Background. Out of the various attempts to mitigate racial bias, the exposure to positive exemplars of the cultural outgroup and to negative exemplars of the cultural ingroup, as well as an increased identification with the cultural outgroup have shown promising results. However, as some of the interventions presented above rely on the interaction with human interlocutors, these solutions are costly and dependent on the confederates' performance. Furthermore, even though the interventions have been shown to lead to a reduction of implicit racial bias, it is not clear how long this effect persists within the participants. In order for a long term reduction of racial bias to develop, repeated interventions might be necessary, which entails logistic challenges and an even greater dependence on the confederates' consistent performance. With IVAs being able to act as social interaction partners in cultural scenarios, they represent a promising tool in the aim of mitigating racial bias. Relying on successful techniques derived from social psychology, they could constitute a cost-effective solution to reduce racial bias - enabling realistic multicultural social situations while profiting from the adaptability and scalability of digital systems. With immersive technologies such as VR, the effect of such an intervention is supposedly even stronger.

3 RELATED WORK WITH IVAS

IVAs have been shown to be successfully used as social interaction partners, since they mostly trigger perceptions and behaviours congruent with those observed in human-human interaction (e.g., [43, 51, 52]). This has also led to a growing number of research efforts that implement IVAs as interaction partners in cultural scenarios.

Rossen et al. conducted a study in which they evaluated whether IVAs can elicit racial bias congruent with the participants' real world racial bias [49]. They assessed the participants' empathy towards a White and a Black IVA and the participants' implicit real-world racial bias towards Black people using the Implicit Association Test. Their results reveal a significant correlation between the participants' empathy towards the White IVA and their implicit racial bias towards Black people. This suggests that IVAs are suitable for triggering racial biases consistent with real world racial biases.

3.1 Implementation and Perception of Mixed-Cultural IVAs

Exploring the design and subsequent perception of enculturated IVAs, numerous studies have investigated the role of different external culture-specific markers, such as appearance (e.g., [20, 56]), non-verbal (e.g., [28, 29, 31, 33]) and verbal behaviour (e.g., [39, 41]). However, research investigating mixed-cultural backgrounds in IVAs is rare.

Adapting the appearance, the verbal and non-verbal behaviour of an IVA to resemble an Arabic-German mixed-cultural background, Lugrin et al. were able to show that participants adapt their own verbal and non-verbal behaviour when interacting with the IVA in a way that is congruent with human-human interaction [32].

With speech being one of the most important markers of mixed-cultural individuals [1], there have been several studies specifically investigating non-native speech with IVAs.

Obremski et al. conducted a series of perception-based studies to examine the role of non-native speech in the development of mixed-cultural IVAs. Conducting two studies on the perception of IVAs making grammatical mistakes typical for non-native speakers in either English or German, they identified thresholds of language proficiency below which IVAs are consistently perceived as non-native speakers of English or German [41, 42]. For the English speaking IVA they found that when it omits at least 50% of the articles and prepositions or when it uses the infinitive form of 50% of the conjugated verbs in a text, it is consistently perceived as a non-native speaker of English. For the German speaking IVA they found that when it misplaces at least 10% of the words or when it uses the infinitive form of 25% of the conjugated verbs in a text, it is consistently perceived as a non-native speaker of German.

Investigating another marker of mixed-cultural speech with IVAs, Obremski et al. compared the perception of non-native accents in English and German using synthetic and natural speech [38, 39]. For the English speaking IVA their results revealed different effects of synthetic and natural speech, depending on the non-native accent used. For the German speaking IVA they were able to show that the synthetically generated non-native accents performed equally good to the natural non-native accents in almost all aspects, endorsing the use of synthetic speech for the implementation of German speaking mixed-cultural IVAs. In another study they were able to further support these findings with Turkish-German mixed-cultural participants showing similar perception of natural and synthetic speech recordings that resembled their own non-native accent [40].

Using natural speech, Dehghani et al. [11] and Khooshabeh et al. [27] were able to show, that mixed-cultural participants shift their interpretative lens in a social situation, when the mixed-cultural IVA they interacted with spoke with a non-native accent that resembled their own mixed-cultural background. Investigating this effect for an Iranian-American and Chinese-American accent, they also were able to show that the IVA was perceived as being more Middle Eastern or more Asian, respectively.

3.2 Intercultural Training and Interventions

The research efforts reported above have shown that IVAs are not only capable of simulating social situations in general, but also of representing specific (mixed-)cultural backgrounds. This has

facilitated different approaches of intercultural training and interventions to raise cultural awareness and increase intercultural empathy. Implementing IVAs in interactive scenarios enables users to actively engage in different contexts rather than passively observing information, which is supposed to increase the ability to later recall the learned information [8].

While some works have attempted to raise general cultural awareness by implementing synthetic cultures in IVAs [2–4, 10], others have targeted the implementation of real-world cultures to train for specific situations.

One system that uses enculturated IVAs to teach the successful interaction with people from the Arabic culture is the Tactical Language and Culture Training System [22, 23, 60]. Apart from courses that teach about the Iraqi and Afghan language and culture, the system also includes courses targeting the French culture [24]. While the system is widely used by military service members, it is also used by business executives to learn for example about the Chinese culture [25]. While the systems presented above focus on intercultural communication and the general raising of cultural awareness, there is, to the best knowledge of the authors, no system that utilises enculturated IVAs to reduce racial biases towards foreign cultures. Several attempts to reduce racial bias have been made in the adjacent research field of avatars [18, 44], e.g., by giving White users a Black avatar. However, these attempts are limited in that they solely rely on manipulating visual culture-specific features in the user's avatar while neglecting other culture-specific parameters as well as the interactive component of intercultural communication. With IVAs being capable of incorporating all of these parameters in a realistic cultural scenario they are predestined to contribute to the goal of reducing real-world racial bias in a controllable and cost-effective manner.

Contribution. This paper presents the implementation and subsequent evaluation of a prototype that is used to create an intuitive interaction with multiple IVAs in VR. Aiming at reducing implicit and explicit racial bias, one of the IVAs was implemented to represent an East African-German mixed-cultural background, while the other was implemented to represent a German mono-cultural background. Based on previous research on the reduction of racial bias, the East African-German mixed-cultural IVA was designed to serve as a positive outgroup exemplar by appearing sympathetic and to trigger ingroup feelings and empathy within the participants. The German mono-cultural IVA was designed to serve as a negative ingroup exemplar by appearing antipathetic and impolite.

Based on the research presented in the previous sections, we formulate the following hypotheses:

H1: Participants' implicit racial bias is lower after the intervention than before the intervention.

H2: Participants' explicit racial bias is lower after the intervention than before the intervention.

4 IMPLEMENTATION

A scenario was implemented to enable the participants' interaction with multiple IVAs in a virtual environment using the game engine Unity (Version 2021.3.6f1). The requirements for the system were to enable a natural interaction with three seemingly autonomous IVAs in virtual reality and to give the experimenter the ability

to control the conversation without the participants being aware of it. Targeted at university students, the narrative was that the participants would take part in a pub quiz with two fellow students, moderated by a bartender.

One IVA resembled a student with an East African-German mixed-cultural background and triggered feelings of sympathy and of a social ingroup within the participants. The other IVA represented a student with a German mono-cultural background and was supposed to trigger antipathy and the feeling of belonging to a social outgroup within the participants. The third IVA resembled a neutral character that acts as a bartender in the virtual scenario. The following section describes the implementation of the VR-environment, the three IVAs, the interactive dialogue, and the Wizard-of-Oz interface respectively.

4.1 Virtual Environment

The virtual environment was designed to resemble a pub with several tables and chairs as well as a bar in the middle of the room. The participant was seated at a round table that matched the table installed in the VR laboratory to provide a congruent experience of visual stimuli and haptic feedback. The mixed-cultural IVA and the mono-cultural IVA were seated at the opposite side of the table while the neutral IVA was standing behind the bar out of the participants natural viewing angle. A screen was placed above the IVAs' heads for displaying relevant information during the quiz game (see Figure 1 B)). To create the illusion of a crowded pub without having to implement the behaviour of many additional IVAs that might distract the participants, the table the participant sat at was separated from the other tables by partitions. In the background an ambient recording of a crowded bar was played to generate the illusion of other guests sitting behind the partitions.

4.2 Intelligent Virtual Agents

4.2.1 Appearance. All three IVAs were designed using Character Generator by Autodesk[1]. The mixed-cultural IVA was designed as a young Black woman with a typical East African appearance. This was achieved by choosing a dark eye colour, a long multicoloured dress and a headscarf. The mono-cultural IVA had to resemble a German background and was designed as a young female character with short blond hair and blue eyes, wearing a black top and light grey pants. Figure 1 B) shows the two IVAs that were used for the intervention. The IVAs were designed as female characters to minimise gender effects, since the majority of the participants were expected to be of female gender. The neutral IVA that served as a bartender was designed as a young male character with short blond hair wearing a grey shirt and green pants.

4.2.2 Verbal Behaviour. The IVAs' verbal behaviour was generated using the text-to-speech engine Amazon Polly[2]. The speech output for all IVAs was generated using the German text-to-speech engine provided by Amazon Polly. For the antipathetic German monocultural IVA the female voice named 'Marlene' was selected, and for the neutral IVA the male voice named 'Hans' was selected. All of the two IVAs' text modules were then run through the engine without

---

[1]https://charactergenerator.autodesk.com
[2]https://aws.amazon.com/de/polly/

Characters in German text	Standard German pronunciation	Ethiopian-German pronunciation
ich	iç	iʂ
r	r	ɽ
z	t͡s	s
habe	ˈhaːbə	ap

Table 1: Overview of the pronunciation rules applied to the German text to generate the Ethiopian accent, based on the International Phonetic Alphabet

Standard German	Ethiopian-German mixed-cultural speech
Meine Mutter macht gerade eine Ausbildung zur Krankenschwester und sie meinte letztens, dass die Niere ja aussehen würde wie eine Bohne.	Meine Mutter <u>machen</u> gerade eine Ausbildung **sur** Krankenschwester und sie meinte letztens, dass die Niere ja aussehen würde wie eine Bohne.

Table 2: Example of how the German text was altered to match the sympathetic IVA's mixed-cultural background. The underlined word resembles an infinitive mistake and the word in bold resembles modified pronunciation.

Standard English	Ethiopian-English mixed-cultural speech
My mother is currently training to be a nurse and she said the other day that the kidney would look like a bean.	My mother currently <u>train</u> to be a nurse and she said the other day that the kidney would look like a bean.

Table 3: English translation of Table 2. The underlined word resembles an infinitive mistake. Due to the specific phonetic characteristics of the German language, the modified pronunciation could not be transferred to the English language.

any further modification to create synthetic speech with standard German pronunciation. For the sympathetic mixed-cultural IVA the voice named 'Vicky' was selected. Here, several adaptations had to be made for the speech to match the IVA's intended East African-German mixed-cultural background. Based on research on the characteristics of non-native speech [1, 47] an East African non-native accent and typical grammatical mistakes were induced.

Previous research conducted by Obremski et al. suggests that in the German language, IVAs with non-native accents are perceived similar when using synthetic speech compared to natural speech [39]. Therefore the mixed-cultural IVA's speech was also generated using Amazon Polly. The text-to-speech engine did not provide the function to automatically generate synthetic German speech with an accent from an East African language, which is why the accent had to be implemented manually. For this, four speech recordings of two female mixed-cultural Ethiopian-German speakers were analysed by German native speakers with regard to anomalies in pronunciation. Table 1 shows the pronunciation rules that were derived from the recordings and applied to the German text to achieve an Ethiopian accented German pronunciation. Either phonetic spelling or the Speech Synthesis Markup Language (SSML) and the respective tags to enable specific pronunciation based on the International Phonetic Alphabet[3] was used.

Regarding grammatical mistakes previous research has shown that by changing 25% of the conjugated verbs back to their infinitive form or by misplacing 10% of the words in German speech, the respective IVA is perceived as a non-native speaker of German [41]. Since changing verbs to their infinitive form is less disruptive to the sentence structure, this alteration was applied to the mixed-cultural IVA's speech. Additionally, the general disregard of grammatical cases in possessive pronouns (e.g., *mine* or *yours*), which was derived as a typical mistake of Ethiopian-German speakers, was induced into the text of the mixed-cultural IVA. Table 2 shows an example of the modifications applied to the German text to achieve non-native speech with an Ethiopian accent, Table 3 shows the same example translated to English.

For all IVAs and their respective verbal behaviour, several .mp3 files were generated and imported into Unity.

4.2.3 Non-Verbal Behaviour. The non-verbal behaviour for the three IVAs was implemented by integrating generic gestures from

Mixamo[4] to visually support the IVAs' verbal behaviour. The animations did mostly resemble beat gestures, except for one gesture of the East African-German mixed-cultural IVA pulling out its smartphone to submit the answers during the pub quiz. The IVAs' facial expressions and lip movements were generated using the Unity plugin SALSA LipSync Suite[5]. While the lip movements were generated automatically based on the plugin's approximation algorithm, the IVAs' facial expressions (such as smiles or raised eyebrows) were explicitly triggered to match the IVAs' verbal behaviour. To emphasise the mixed-cultural IVA's sympathetic traits, it was implemented to smile more often than the German mono-cultural IVA.

4.3 Interactive Scenario

To ensure a predictable and seamless conversation between the participants and the two IVAs, the conversation was designed to always be controlled by the IVAs, involving the participant by addressing him or her directly.

The interactive scenario was designed to take place in a pub during a party organised by the participants' university. The conversation started with the two IVAs introducing themselves and making small talk with the participant about their respective field of studies. The mixed-cultural IVA told the participant that it was an East African immigrant to clarify its intended mixed-cultural background. To trigger the impression of a social ingroup, the mixed-cultural IVA first asked the participant about his or her field

---

[3]https://www.internationalphoneticalphabet.org/

[4]https://mixamo.com
[5]https://assetstore.unity.com/packages/tools/animation/salsa-lipsync-suite-148442

of studies and subsequently stated that it was studying the same. Since the pool of participants was limited to students of three different bachelor programs, these three options were included as potential answers of the mixed-cultural IVA. To obtain the impression of the German mono-cultural IVA to be antipathetic and rude, the IVA expressed lack of appreciation for the aforementioned bachelor program and stated it was studying business administration for a better career perspective. The mixed-cultural IVA reacted in a kind manner and expressed its interest in the German mono-cultural IVA's studies to further foster its sympathetic impression.

In the next step, the IVA that acted as a bartender announced the start of the pub quiz. The pub quiz consisted of three questions that were asked by the bartender and displayed on the screen above the two IVAs' heads. For each question four possible answers were also displayed on the screen, while answers should be submitted via a fictional smartphone app. During the quiz, the mono-cultural IVA consistently insisted on being right while giving the wrong answers and refusing every suggestion made by the mixed-cultural IVA and the participant. The mixed-cultural IVA on the other hand kept acting kind and diplomatic. After the two IVAs and the participants came to a decision, the mixed-cultural IVA pulled out its smartphone to submit the answer, to further foster its supportive actions for the team. The interaction ended after the submission of the answer to the last question and the IVA that acted as a bartender declared the end of the game.

4.4 Wizard-of-Oz Interface

A Wizard-of-Oz interface was implemented to enable the experimenter to control the conversation. The interface consisted of multiple buttons, each of which resembled a specific verbal and non-verbal behaviour of the IVAs. After each time the participant took part in the conversation, the experimenter reacted by triggering the suitable response of the IVAs, oftentimes choosing between multiple options. To react to unforeseen answers from the participant, additional buttons were implemented that triggered generic reactions from the IVAs, such as them saying *yes* or *never mind*. The Wizard-of-Oz interface was tested multiple times with different people to ensure that there are suitable reactions for the majority of the participants' possible answers. It was also iteratively optimised to ensure an intuitive use for the experimenter to minimise the pauses between the participants' reactions and the IVAs' responses. Figure 1 A) shows the experimenter using the interface.

5 USER STUDY

To test whether implicit and explicit racial biases can be reduced by interacting with a sympathetic East African-German mixed-cultural IVA and an antipathetic German mono-cultural IVA, a user study was conducted. The study was designed as a repeated-measures VR experiment in which every participant experienced the same treatment. The HTC VIVE Pro [6] was used as a VR headset to display the virtual environment. One of the experimenters was Black and made sure that the minority of Blacks in Germany was adequately represented. The study was approved by the local ethics committee (reference number: 230622).

---

[6]https://www.vive.com/de/product/vive-pro/

5.1 Participants

In total, N = 32 participants between 20 and 26 years took part in the VR study (M_{age} = 22.37, SD_{age} = 1.81). 23 of the participants were female, eight were male and one was of diverse gender. Furthermore, all participants reported to be of Caucasian ethnicity. Two participants reported a mixed ethnicity, with one naming Black and one naming Latin American in addition to the Caucasian ethnicity. Every participant received partial course credit for participating in the study.

5.2 Measures

5.2.1 Manipulation Check. To assess whether the manipulation in our intervention was successful, several measures were taken.

Warmth and Competence. The two IVAs' perceived warmth and competence was measured, using the questionnaire by Nass et al. [37]. The questionnaire consists of six items for each of the two dimensions. Participants were asked to indicate the extent to which they considered various adjectives to apply to each IVA, on a 10-point Likert scale. The dimension *warmth* included adjectives such as *warm*, *friendly*, or *polite* whereas the dimension *competence* included adjectives such as *useful*, *analytic*, or *competent*.

Sympathy. Assessing whether the East African-German mixed-cultural IVA was perceived as more sympathetic than the German mono-cultural IVA, the participants were asked to indicate their agreement with the following sentence on a 7-point Likert scale: *'This IVA was sympathetic'*.

Empathy. Evaluating the participants' empathy towards the two IVAs, the empathy scale by Rosenthal-von der Pütten et al. [48] was used for both IVAs. Participants were asked to indicate their agreement with in total five statements like *'This IVA did not feel anything'* on a 5-point Likert scale.

5.2.2 Racial Bias.

Implicit racial bias. The implicit racial bias was measured twice, once before and once after the intervention, using the Implicit Association Test (IAT). The open source tool FreeIAT was used to conduct the test [35]. The IAT consisted of five rounds in total, out of which the third and the fifth round were the two test trails while the others were learning trials. The IAT was .modified to include photographs of five Caucasian and five East African women with neutral facial expressions (see Figure 2), as well as nine negative (e.g., *horrible*, *agony*, *failure*) and nine positive (e.g., *joy*, *wonderful*, *laughter*) words. As the final score of the participants' racial bias, the Greenwald, Nosek, and Banaji overall IAT score [17] was calculated which results in scores between -2 and 2. In the context of this study, positive values indicate a higher association of negative words with East African women, while negative values indicate a higher association of negative words with Caucasian women. A score of 0 would represent no specific preference for either of the two and as such no bias towards either group.

Explicit racial bias. To assess the participants' explicit racial bias, the blatant and subtle prejudice scale by Meertens and Pettigrew was used [36]. These scales were also assessed twice, once before

Figure 2: Collection of the visual stimuli presented in the IAT. The upper row resembles photographs of Caucasian women, the lower row resembles photographs of East African women. Faces were blurred to preserve anonymity.

and once after the interaction. The scales were modified to explicitly inquire the opinion on East African immigrants in Germany. Participants were asked to indicate their agreement with statements such as *'East African refugees have jobs that Germans should have.'* on a four-point Likert scale.

5.3 Procedure

After filling in the consent forms, the participants were asked to fill in the first part of the questionnaire which contained the pre-measures for the explicit racial bias towards East African immigrants. After that, the baseline of the participants' implicit racial bias was assessed by using the IAT. The participants were then seated at a round table in the laboratory where the experimenter handed them the VR-Headset. The experimenter started the virtual scenario and controlled the Wizard-of-Oz interface with a silent mouse to preserve the illusion of interacting with autonomous IVAs. The intervention in VR took between six and ten minutes, depending on the conversational behaviour of the participant. After the intervention the participants filled in the second part of the questionnaire, containing the post-measures for the explicit racial bias, the questionnaires for the manipulation check, the demographic questionnaires and the post-measures for the implicit racial bias in form of a second IAT.

5.4 Results

All calculations were conducted using SPSS Statistics (Version 21). In total, six t-tests for paired samples were calculated. To correct for the higher probability of alpha errors due to multiple t-tests, the alpha of $p = .050$ was corrected to $p = .008$ to indicate significance.

5.4.1 Manipulation Check. The mixed-cultural IVA was rated significantly higher in warmth ($M = 9.39$, $SD = .67$) than the mono-cultural IVA ($M = 1.70$, $SD = .66$), $t(31) = 44.17$, $p < .001$, $d = 7.81$.

The mixed-cultural IVA was also rated significantly higher ($M = 8.54$, $SD = 1.27$) in competence than the mono-cultural IVA ($M = 3.53$, $SD = 1.82$), $t(31) = 11.75$, $p < .001$, $d = 2.08$. In terms of sympathy, participants rated the mixed-cultural IVA to be significantly more sympathetic ($M = 6.69$, $SD = .64$) than the mono-cultural IVA ($M = 1.25$, $SD = .57$), $t(31) = .31.33$, $p < .001$, $d = 5.54$. Participants also reported significantly more empathy towards the mixed-cultural IVA ($M = 3.81$, $SD = .54$) than towards the mono-cultural IVA ($M = 2.36$, $SD = .64$), $t(31) = 9.14$, $p < .001$, $d = 1.61$.

5.4.2 Racial Bias. Based on the values of the IAT, the implicit racial bias of the participants was significantly lower after the interaction ($M = -.04$, $SD = .45$) than before the interaction ($M = .37$, $SD = .42$), $t(31) = -4.85$, $p < .001$, $d = .86$. Figure 3 shows the results of the pre- and post-intervention measures of implicit racial bias.
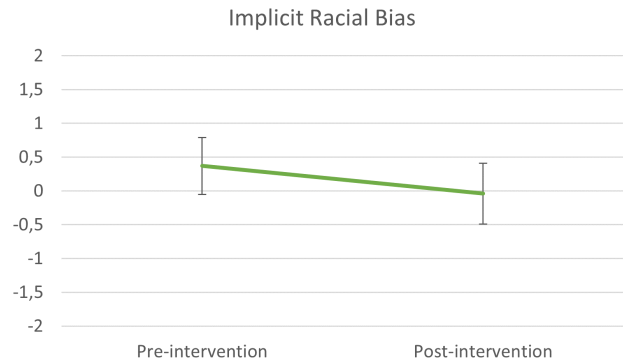

Figure 3: Results of the pre- and post-intervention measures of implicit racial bias. The error bars show the standard deviation.

The explicit bias also decreased, showing significantly lower values after ($M = 1.83$, $SD = .36$) than before the intervention ($M = 1.92$, $SD = .29$), $t(31) = 3.41$, $p = .002$, $d = .60$. Figure 4 shows the results of the pre- and post-intervention measures of explicit racial bias.
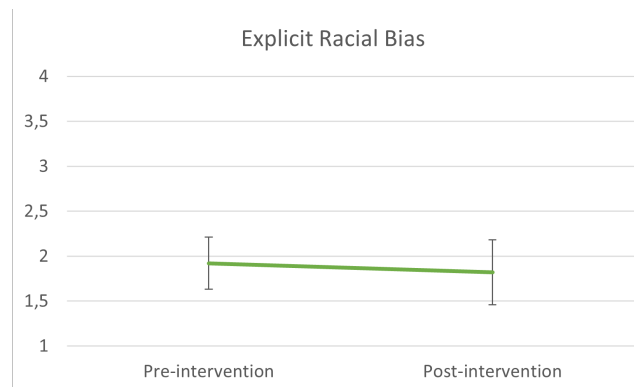

Figure 4: Results of the pre- and post-intervention measures of explicit racial bias. The error bars show the standard deviation.

6 DISCUSSION

The variables assessed for the manipulation check revealed that the East African mixed-cultural IVA was perceived as significantly warmer and more competent than the German mono-cultural IVA. As people usually rate members of their own ingroup high on these two dimensions [59], it can be assumed that despite the IVA

being a member of a cultural outgroup, its simulated shared interests with the participants were sufficient to evoke the feeling of a social ingroup. This was also observed in the participants' significantly higher empathy towards the East African mixed-cultural IVA compared to the German mono-cultural IVA which indicates more perceived similarities with the first [63]. Finally, the significantly higher rating in sympathy for the East African mixed-cultural IVA shows that the manipulation of creating one sympathetic mixed-cultural IVA and one antipathetic mono-cultural IVA was successful. As predicted by H1, the intervention with the two IVAs led to a significant decrease of the participants' implicit racial bias. Hence, H1 can be accepted. This shows that the techniques of exposing the participant to a positive outgroup and a negative ingroup exemplar [26, 30] combined with the increased identification with the cultural outgroup via shared interests [19, 62] can be used with IVAs for changing implicit attitudes within human interlocutors during a short intervention in VR. Despite the fact that participants' implicit bias was relatively low to begin with, it still decreased significantly through the intervention.

H2 predicted that the intervention would lead to a significant decrease of the participants' explicit racial bias. With the participants showing significantly lower scores of explicit racial bias after the intervention compared to before the intervention, H2 can also be accepted. Again, with the pre-intervention scores for explicit racial bias being at $M = 1.92$ on a four point Likert scale, the participants' explicit bias was low to begin with. The high level of education and the young age of the participants might have affected the low pre-intervention biases, and might be a sign of a generally high cultural awareness. In addition, social desirability [57] and the self-perception of being free of prejudice may have led participants to agree less with explicit statements against foreigners, even if they did so subconsciously [12]. It would therefore be interesting to investigate whether the intervention leads to similar effects with participants that have a higher baseline in terms of implicit and explicit racial bias.

Furthermore it is yet to be investigated how long the reduced racial bias persists within the participants. Since this is a general concern with effects that are based on intergroup contact [21], it should be further investigated in longitudinal studies.

7 CONCLUSION AND FUTURE WORK

This paper presented the implementation of a scenario that was designed to enable the reduction of racial bias through a Wizard-of-Oz controlled intervention with multiple IVAs. A repeated-measures VR-based user study revealed that the intervention led to a significant reduction of both, implicit and explicit bias towards East African immigrants. These results constitute important implications for the use of IVAs in cultural training applications and set the base for further research on the reduction of racial bias using IVAs.

In our future work we will investigate the persistence of the reduced racial bias in longitudinal studies. By assessing the participants' racial bias again one week after the intervention we will evaluate if a single exposure is sufficient or if multiple exposures are needed to achieve a long-term reduction of racial bias. Furthermore, we will test if the reduction of racial bias transfers to cultures other than the East African one presented in the current paper.

And, to test whether the success of the intervention correlates with the participants' openness towards foreign cultures, we will test whether the reduction of racial bias also works when the sample has a higher baseline in both implicit and explicit racial bias.

REFERENCES

[1] Matthew P. Aylett, Leigh Clark, Benjamin R. Cowan, and Ilaria Torre. 2021. *Building and Designing Expressive Speech Synthesis* (1 ed.). Association for Computing Machinery, New York, NY, USA, 173–212. https://doi.org/10.1145/3477322.3477329

[2] R. Aylett, L. Hall, S. Tazzyman, B. Endrass, E. André, C. Ritter, A. Nazir, A. Paiva, G.J. Hofstede, and A. Kappas. 2014. Werewolves, Cheats, and Cultural Sensitivity. In *Int. Conf. on Autonomous Agents and Multiagent Systems (AAMAS 2014)*. IFAAMAS, 1085–1092.

[3] Ruth Aylett, Michael Kriegel, Mei Yii Lim, Joao Dias, Karin Leichtenstern, Wan Ching Ho, and Paola Rizzo. 2009. ORIENT: interactive agents for stage-based role-play. In *Proceedings of The 8th International Conference on Autonomous Agents and Multiagent Systems-Volume 2*. 1371–1372.

[4] Ruth Aylett, Natalie Vannini, Elisabeth Andre, Ana Paiva, Sibylle Enz, and Lynne Hall. 2009. *But that was in another country: agents and intercultural empathy.* International Foundation for Autonomous Agents and Multiagent Systems.

[5] Tiffany N Brannon and Gregory M Walton. 2013. Enacting cultural interests: How intergroup contact reduces prejudice by sparking interest in an out-group's culture. *Psychological Science* 24, 10 (2013), 1947–1957.

[6] Marilynn B. Brewer. 1999. The Psychology of Prejudice: Ingroup Love and Outgroup Hate? *Journal of Social Issues* 55, 3 (1999), 429–444. https://doi.org/10.1111/0022-4537.00126 arXiv:https://spssi.onlinelibrary.wiley.com/doi/pdf/10.1111/0022-4537.00126

[7] Statistisches Bundesamt. 2022. Weltflüchtlingstag: Mindestens 3,3 Millionen Geflüchtete und Vertriebene in Deutschland. https://www.destatis.de/DE/Presse/Pressemitteilungen/2022/06/PD22_N036_12411.html

[8] Chris Christou. 2010. Virtual reality in education. In *Affective, interactive and cognitive methods for e-learning design: creating an optimal education experience.* IGI Global, 228–243.

[9] Corey Columb and E Ashby Plant. 2011. Revisiting the Obama effect: Exposure to Obama reduces implicit prejudice. *Journal of Experimental Social Psychology* 47, 2 (2011), 499–501.

[10] Nick Degens, Gert Jan Hofstede, Samuel Mascarenhas, André Silva, Ana Paiva, Felix Kistler, Elisabeth André, Aleksandra Swiderska, Eva Krumhuber, Arvid Kappas, et al. 2013. Traveller–intercultural training with intelligent agents for young adults. (2013).

[11] Morteza Dehghani, Peter Khooshabeh, Lixing Huang, Lia Oganesyan, and Jonathan Gratch. 2011. Cultural frame-switching using accented spoken language by a virtual character. In *Workshop on Culturally Motivated Virtual Characters, IVA.*

[12] John F Dovidio, Kerry Kawakami, and Samuel L Gaertner. 2000. Reducing contemporary prejudice: Combating explicit and implicit bias at the individual and intergroup level. *Reducing prejudice and discrimination* (2000), 137–163.

[13] John F Dovidio, Kerry Kawakami, and Samuel L Gaertner. 2002. Implicit and explicit prejudice and interracial interaction. *Journal of personality and social psychology* 82, 1 (2002), 62.

[14] Chloë FitzGerald, Angela Martin, Delphine Berner, and Samia Hurst. 2019. Interventions designed to reduce implicit prejudices and implicit stereotypes in real world contexts: a systematic review. *BMC psychology* 7, 1 (2019), 1–12.

[15] Feng Fu, Corina E Tarnita, Nicholas A Christakis, Long Wang, David G Rand, and Martin A Nowak. 2012. Evolution of in-group favoritism. *Scientific reports* 2, 1 (2012), 1–6.

[16] Anthony G Greenwald, Debbie E McGhee, and Jordan LK Schwartz. 1998. Measuring individual differences in implicit cognition: the implicit association test. *Journal of personality and social psychology* 74, 6 (1998), 1464.

[17] Anthony G Greenwald, Brian A Nosek, and Mahzarin R Banaji. 2003. Understanding and using the implicit association test: I. An improved scoring algorithm. *Journal of personality and social psychology* 85, 2 (2003), 197.

[18] Victoria Groom, Jeremy N Bailenson, and Clifford Nass. 2009. The influence of racial embodiment on racial bias in immersive virtual environments. *Social Influence* 4, 3 (2009), 231–248.

[19] Natalie R Hall, Richard J Crisp, and Mein-woei Suen. 2009. Reducing implicit prejudice by blurring intergroup boundaries. *Basic and Applied Social Psychology* 31, 3 (2009), 244–254.

[20] Béatrice S Hasler, Gilad Hirschberger, Tal Shani-Sherman, and Doron A Friedman. 2014. Virtual peacemakers: Mimicry increases empathy in simulated contact with virtual outgroup members. *Cyberpsychology, Behavior, and Social Networking* 17, 12 (2014), 766–771.

[21] Miles Hewstone, Simon Lolliot, Hermann Swart, Elissa Myers, Alberto Voci, Ananthi Al Ramiah, and Ed Cairns. 2014. Intergroup contact and intergroup

conflict. *Peace and Conflict: Journal of Peace Psychology* 20, 1 (2014), 39.

[22] W Lewis Johnson, Stacy Marsella, and Hannes Vilhjalmsson. 2004. The DAR-WARS tactical language training system. In *Proceedings of I/ITSEC*.

[23] W Lewis Johnson and Andre Valente. 2008. Collaborative authoring of serious games for language and culture. In *Proceedings of SimTecT*, Vol. 2008.

[24] W Lewis Johnson and Andre Valente. 2009. Tactical language and culture training systems: Using AI to teach foreign languages and cultures. *AI magazine* 30, 2 (2009), 72–72.

[25] W Lewis Johnson and Sara Behani Zaker. 2012. The power of social simulation for Chinese language teaching. *Proceedings of TCLT7* (2012).

[26] Jennifer A Joy-Gaba and Brian A Nosek. 2010. The surprisingly limited malleability of implicit racial evaluations. *Social Psychology* 41, 3 (2010), 137.

[27] Peter Khooshabeh, Morteza Dehghani, Angela Nazarian, and Jonathan Gratch. 2017. The cultural influence model: when accented natural language spoken by virtual characters matters. *AI & SOCIETY* 32 (2017), 9–16.

[28] Tomoko Koda and Zsofia Ruttkay. 2017. Eloquence of eyes and mouth of virtual agents: cultural study of facial expression perception. *AI & society* 32, 1 (2017), 17–24.

[29] Tomoko Koda, Zsofia Ruttkay, Yuka Nakagawa, and Kyota Tabuchi. 2010. Cross-cultural study on facial regions as cues to recognize emotions of virtual agents. In *Culture and Computing*. Springer, 16–27.

[30] Calvin K Lai, Maddalena Marini, Steven A Lehr, Carlo Cerruti, Jiyun-Elizabeth L Shin, Jennifer A Joy-Gaba, Arnold K Ho, Bethany A Teachman, Sean P Wojcik, Spassena P Koleva, et al. 2014. Reducing implicit racial preferences: I. A comparative investigation of 17 interventions. *Journal of Experimental Psychology: General* 143, 4 (2014), 1765.

[31] Birgit Lugrin, Andrea Bartl, Hendrik Striepe, Jennifer Lax, and Takashi Toriizuka. 2018. Do I act familiar? Investigating the similarity-attraction principle on culture-specific communicative behaviour for social robots. In *2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2033–2039.

[32] B. Lugrin, B. Eckstein, K. Bergmann, and C. Heindl. 2018. Adapted Foreigner-directed Communication Towards Virtual Agents. In *18th Int. Conf. on Intelligent Virtual Agents* (Sydney, NSW, Australia) *(IVA '18)*. ACM, 59–64. https://doi.org/10.1145/3267851.3267859

[33] Birgit Lugrin, Julian Frommel, and Elisabeth André. 2018. Combining a data-driven and a theory-based approach to generate culture-dependent behaviours for virtual characters. In *Advances in Culturally-Aware Intelligent Systems and in Cross-Cultural Psychological Studies*. Springer, 111–142.

[34] Joshua A McGrane and Fiona A White. 2007. Differences in Anglo and Asian Australians' explicit and implicit prejudice and the attenuation of their implicit in-group bias. *Asian Journal of Social Psychology* 10, 3 (2007), 204–210.

[35] Adam W Meade. 2009. FreeIAT: An open-source program to administer the Implicit Association Test. *Applied psychological measurement* (2009).

[36] Roel W Meertens and Thomas F Pettigrew. 1997. Is subtle prejudice really prejudice? *The Public Opinion Quarterly* 61, 1 (1997), 54–71.

[37] Clifford Nass, Youngme Moon, and Paul Carney. 1999. Are people polite to computers? Responses to computer-based interviewing systems 1. *Journal of applied social psychology* 29, 5 (1999), 1093–1109.

[38] David Obremski, Helena Hering, Paula Friedrich, and Birgit Lugrin. 2022. Mixed-Cultural Speech for Intelligent Virtual Agents - the Impact of Different Non-Native Accents Using Natural or Synthetic Speech in the English Language. (2022). https://doi.org/10.1145/3527188.3561921 (in press).

[39] David Obremski, Helena Babette Hering, Paula Friedrich, and Birgit Lugrin. 2022. Exploratory Study on the Perception of Intelligent Virtual Agents With Non-Native Accents Using Synthetic and Natural Speech in German. *INTERNATIONAL CONFERENCE ON MULTIMODAL INTERACTION* (2022). https://doi.org/10.1145/3536221.3556608 (in press).

[40] David Obremski and Birgit Lugrin. 2022. Mixed-Cultural Speech for Mixed-Cultural Users - Natural vs. Synthetic Speech for Virtual Agents. (2022). https://doi.org/10.1145/3527188.3563910 (in press).

[41] David Obremski, Jean-Luc Lugrin, Philipp Schaper, and Birgit Lugrin. 2019. Non-Native Speaker Generation and Perception for Mixed-Cultural Settings. In *Proceedings of the 19th ACM International Conference on Intelligent Virtual Agents* (Paris, France) *(IVA '19)*. ACM, New York, NY, USA, 105–107. https://doi.org/10.1145/3308532.3329427

[42] David Obremski, Jean-Luc Lugrin, Philipp Schaper, and Birgit Lugrin. 2021. Non-native speaker perception of Intelligent Virtual Agents in two languages: the impact of amount and type of grammatical mistakes. *Journal on Multimodal User Interfaces* 15, 2 (2021), 229–238.

[43] David Obremski, Alicia L Schäfer, Benjamin P Lange, Birgit Lugrin, Elisabeth Ganal, Laura Witt, Tania R Nuñez, Sascha Schwarz, and Frank Schwab. 2021. Put that Away and Talk to Me-the Effects of Smartphone induced Ostracism while Interacting with an Intelligent Virtual Agent. In *Proceedings of the 9th International Conference on Human-Agent Interaction*. 428–432.

[44] Tabitha C Peck, Sofia Seinfeld, Salvatore M Aglioti, and Mel Slater. 2013. Putting yourself in the skin of a black avatar reduces implicit racial bias. *Consciousness and cognition* 22, 3 (2013), 779–787.

[45] David G Rand, Thomas Pfeiffer, Anna Dreber, Rachel W Sheketoff, Nils C Wern-erfelt, and Yochai Benkler. 2009. Dynamic remodeling of in-group bias during the 2008 presidential election. *Proceedings of the National Academy of Sciences* 106, 15 (2009), 6187–6191.

[46] United Nations High Commissioner for Refugees. 2022. UNHCR Global Trends 2015. https://www.unhcr.org/statistics/unhcrstats/576408cd7/unhcr-global-trends-2015.html

[47] Jack Richards. 1971. Error Analysis and Second Language Strategies. *The text of an invited lecture given at Indiana University, Bloomington* (1971), 28.

[48] Astrid M Rosenthal-von der Pütten, Nicole C Krämer, Laura Hoffmann, Sabrina Sobieraj, and Sabrina C Eimler. 2013. An experimental study on emotional reactions towards a robot. *International Journal of Social Robotics* 5, 1 (2013), 17–34.

[49] Brent Rossen, Kyle Johnsen, Adeline Deladisma, Scott Lind, and Benjamin Lok. 2008. Virtual humans elicit skin-tone bias consistent with real-world skin-tone biases. In *International Workshop on Intelligent Virtual Agents*. Springer, 237–244.

[50] Myron Rothbart and Oliver P John. 1985. Social categorization and behavioral episodes: A cognitive analysis of the effects of intergroup contact. *Journal of Social Issues* 41, 3 (1985), 81–104.

[51] Tanja Schneeberger, Sofie Ehrhardt, Manuel S Anglet, and Patrick Gebhard. 2019. Would you follow my instructions if I was not human? Examining obedience towards virtual agents. In *2019 8th International Conference on Affective Computing and Intelligent Interaction (ACII)*. IEEE, 1–7.

[52] Tanja Schneeberger, Mirella Scholtes, Bernhard Hilpert, Markus Langer, and Patrick Gebhard. 2019. Can Social Agents elicit Shame as Humans do?. In *2019 8th International Conference on Affective Computing and Intelligent Interaction (ACII)*. IEEE, 164–170.

[53] Richard Sosis and Bradley J Ruffle. 2003. Religious ritual and cooperation: Testing for a relationship on Israeli religious and secular kibbutzim. *Current Anthropology* 44, 5 (2003), 713–722.

[54] Statista. 2021. Größte Herkunftsländer von Flüchtlingen 2021. https://de.statista.com/statistik/daten/studie/186108/umfrage/herkunftslaender-von-fluechtlingen/

[55] Henri Tajfel, Michael G Billig, Robert P Bundy, and Claude Flament. 1971. Social categorization and intergroup behaviour. *European journal of social psychology* 1, 2 (1971), 149–178.

[56] Ron Tamborini, Eric Novotny, Sujay Prabhu, Matthias Hofer, Clare Grall, Brian Klebig, Lindsay S Hahn, Janine Slaker, Rabindra A Ratan, and Gary Bente. 2018. The effect of behavioral synchrony with black or white virtual agents on outgroup trust. *Computers in Human Behavior* 83 (2018), 176–183.

[57] Matilde Tassinari, Matthias Burkard Aulbach, and Inga Jasinskaja-Lahti. 2022. The use of virtual reality in studying prejudice and its reduction: A systematic review. *PloS one* 17, 7 (2022), e0270748.

[58] John C Turner, Rupert J Brown, and Henri Tajfel. 1979. Social comparison and group interest in ingroup favouritism. *European journal of social psychology* 9, 2 (1979), 187–204.

[59] John C Turner, Michael A Hogg, Penelope J Oakes, Stephen D Reicher, and Margaret S Wetherell. 1987. *Rediscovering the social group: A self-categorization theory.* basil Blackwell.

[60] Andre Valente, W Lewis Johnson, and Hannes Högni Vilhjálmsson. 2006. The Tactical Language and Culture Training System: A Demonstration.. In *AAAI*. 1955–1957.

[61] Sam Whitt and Rick K Wilson. 2007. The dictator game, fairness and ethnicity in postwar Bosnia. *American Journal of Political Science* 51, 3 (2007), 655–668.

[62] Anna Woodcock and Margo J Monteith. 2013. Forging links with the self to combat implicit bias. *Group Processes & Intergroup Relations* 16, 4 (2013), 445–461.

[63] Stephen C Wright, Arthur Aron, Tracy McLaughlin-Volpe, and Stacy A Ropp. 1997. The extended contact effect: Knowledge of cross-group friendships and prejudice. *Journal of Personality and Social psychology* 73, 1 (1997), 73.