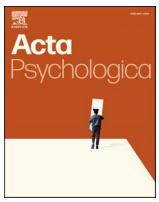
Acta Psychologica 258 (2025) 105129

Contents lists available at ScienceDirect

Acta Psychologica

journal homepage: www.elsevier.com/locate/actpsy



Registered Report Stage I

AI-driven digital humans for *E*-contact: A pre-registered study on reducing intergroup bias with generative artificial intelligence

Manfredi Anna [a,*], Puzella Giulio [a], David Landi [b], Iolanda Iacono [b], Jacopo Michilli [a], Gabbiadini Alessandro [a]

[a] University of Milano Bicocca, Department of Psychology, Mind and Behavior Technological Center, Italy
[b] QuestIT, Via L. Cialfi 23, 53100 Siena, Italy

ARTICLE INFO

Keywords:
E-contact interventions
Prejudice reduction
Artificial intelligence
Intergroup relations
intergroup contact

ABSTRACT

The present pre-registered report explores the potential of AI-driven contact interventions by integrating generative Artificial Intelligence-based artificial humans into *E*-contact paradigms. Grounded in Allport's (1954) Contact Hypothesis and the Dual Identity Model (DIM; Gaertner & Dovidio, 2000), the study examines whether structured interactions with artificial humans can foster positive intergroup attitudes. Following the framework of E-contact interventions (White & Abu-Rayya, 2012), participants ($N = 70$ Caucasian university students) will engage in a within-between mixed-design experiment over three days. They will interact daily with an AI-powered 3D digital assistant representing either an outgroup member (Black avatar) or an ingroup member (White avatar) depending on the experimental condition, with pre- and post-intervention measures of intergroup attitudes. The structured interactions will follow the two-phase design of *E*-contact interventions, initially fostering personal acquaintance, then emphasizing group salience, and finally reinforcing a shared superordinate identity—a process aligned with the principles of DIM to maximize the generalization of positive intergroup attitudes. The virtual assistant will facilitate cooperative activities designed to enhance inclusivity, promote cultural exchange, and maintain subgroup distinctiveness while fostering a common identity. To ensure the effectiveness and coherence of the intervention, the scripted interactions will be pretested through a pilot study before implementation. This research offers a preliminary step toward understanding how artificial intelligence might contribute to enhancing *E*-contact interventions, potentially providing scalable and structured tools for fostering positive intergroup relations and supporting social integration.

In today's increasingly interconnected world, globalization, digital technologies, and growing migration flows have significantly increased interaction opportunities among individuals from diverse social and cultural backgrounds. This unprecedented level of intergroup contact offers substantial benefits, such as promoting intercultural exchange, enhancing mutual understanding, and fostering innovation in multicultural societies. At the same time, it poses challenges, including the persistence and emergence of intergroup prejudice. Consequently, the study of intergroup relations, particularly strategies to reduce bias, remains a central concern in social psychology, highlighting the enduring relevance of classic theoretical frameworks like Allport's (1954) Contact Hypothesis.

Although developed over seven decades ago, the Contact Hypothesis holds theoretical and practical significance in modern contexts (Pettigrew, 1998). Researchers have since expanded upon this model, addressing its limitations and exploring its applications across various domains, including the opportunities afforded by digital technologies. In this regard, society is currently undergoing a technological revolution that many experts (Schwab, 2024) have termed the Fourth Industrial Revolution, marked by the rise of artificial intelligence (AI) technologies. In this regard, recent advancements in virtual environments have demonstrated their potential to promote empathy and reduce bias through immersive interactions (e.g., Bond, 2021; Tassinari et al., 2022a, 2022b). However, empirical studies on integrating AI solutions in prejudice reduction are still scarce. This study addresses this gap by exploring the application of AI-based agents as a facilitator of *E*-contact interventions, building on foundational work by White et al. (2020), thereby expanding the possibilities for intergroup contact (Dovidio

* Corresponding author at: University of Milano-Bicocca, Department of Psychology, Piazza Ateneo Nuovo, 1, 20126 Milano, Italy.
E-mail addresses: a.manfredi17@campus.unimib.it (M. Anna), g.puzella@campus.unimib.it (P. Giulio), d.landi@quest-it.com (D. Landi), iacono@quest-it.com (I. Iacono), j.michilli@campus.unimib.it (J. Michilli), alessandro.gabbiadini@unimib.it (G. Alessandro).

https://doi.org/10.1016/j.actpsy.2025.105129
Received 5 February 2025; Received in revised form 15 May 2025; Accepted 4 June 2025
Available online 12 June 2025

et al., 2017; Pettigrew, 1998) and the E-contact model (White & Abu-Rayya, 2012; White et al., 2020).

1.1. Foundations of intergroup contact theory

Allport's Contact Hypothesis posited that intergroup contact under four optimal conditions—equal status, cooperation, common goals, and institutional support—could significantly reduce prejudice. Subsequent research (Pettigrew & Tropp, 2006), has reinforced these findings, showing that contact is more effective when Allport's conditions are met, but also showed that these conditions don't have to be fulfilled for positive effects to occur. The valence of the contact determines the outcome, whether contact is perceived as positive by the people involved. However, these conditions and the positive valence of contact don't address how positive intergroup attitudes can generalize beyond the immediate interaction. In this context, categorization has emerged as a crucial factor in facilitating the generalization of positive attitude changes to other outgroup members who were not directly involved in the contact.

Several theoretical models have been proposed to explain how intergroup contact can reduce bias, as the *Decategorized Contact Model* (DCM; Brewer & Miller, 1984), the *Common Ingroup Identity Model* (CIIM; Gaertner et al., 1989, 1993; Gaertner & Dovidio, 2014), the *Mutual Intergroup Differentiation Model* (MIDM; Hewstone & Brown, 1986; Vivian et al., 1997; Brown & Hewstone, 2005) and the *Dual Identity model* (DIM; Dovidio et al., 1998; Gaertner et al., 1999; González & Brown, 2003; González & Brown, 2006).

The DCM model (Brewer & Miller, 1984) suggests that simply categorizing people into distinct groups is enough to generate intergroup bias (Tajfel et al., 1971). By reducing the prominence of these group distinctions, the likelihood of making category-based judgments decreases, thereby reducing intergroup bias (Brewer, 1979; Turner, 1981). Since interactions based on group categories often lead to stereotypical perceptions of ingroup and outgroup members (Brown & Turner, 1981), personalizing the contact situation encourages more individualized processing, increasing the potential for disconfirming group stereotypes (Brewer & Miller, 1984). Consequently, intergroup contact should be structured to downplay group categories and promote opportunities for outgroup members to be seen as individuals.

Proponents of the MIDM model (Hewstone & Brown, 1986; Vivian et al., 1997) argued that intergroup contact should be structured to ensure that group identities are not threatened, thus preserving the salience of the original groups. The model encourages groups to emphasize their mutual distinctiveness in the context of equal status interactions and cooperative interdependence so that the members of each group can recognize and appreciate the indispensable contribution of the other (Gaertner et al., 2000). The reduction of intergroup bias is particularly evident in members of the ingroup who interacted with typical-pleasant outgroup members (Brown et al., 1999).

The CIIM model (Gaertner et al., 1989; Gaertner & Dovidio, 2014) suggests that intergroup bias can be reduced by promoting recategorization in the contact situation: strategies should be adopted to recategorize the intergroup situation so that ingroup and outgroup are included in a more inclusive superordinate category. As a result, outgroup members will be seen as part of a common ingroup, thus inhibiting intergroup differentiation and ingroup favoritism based on initial categories (Gaertner & Dovidio, 2014). By altering perceptions of ingroup boundaries, some cognitive and motivational processes that initially contributed to intergroup bias are redirected to foster more positive intergroup relations.

Although all these models effectively reduce prejudice, research has shown that each one has limitations in application. The Dual Identity Model (DIM) (Gaertner et al., 1999; Gaertner & Dovidio, 2000; González & Brown, 2003, 2006) was developed to integrate the strengths of these models while overcoming their respective limitations, particularly regarding the generalization of positive intergroup attitudes beyond the immediate contact situation. Unlike the Decategorization Model (DCM), which fosters personalized contact at the cost of group category salience, DIM preserves subgroup identity alongside a shared superordinate identity. This ensures that positive attitudes are not restricted to individual outgroup members encountered during contact but instead extend to the outgroup as a whole, facilitating broader and more enduring bias reduction (González & Brown, 2003). The Mutual Intergroup Differentiation Model (MIDM), which stresses the distinctiveness of subgroups, also has some disadvantages over the reduction of intergroup prejudice. Indeed, making group membership highly salient during contact can have both positive and negative effects on attitude generalization (Brown et al., 1999; Wilder, 1984). In some cases, greater awareness of intergroup differences may reinforce rather than reduce intergroup boundaries, limiting the possibilities for positive attitudes to extend beyond the immediate interaction. Furthermore, a greater emphasis on the group may be associated with greater intergroup anxiety, which in turn may negatively influence attitudes toward the outgroup (Greenland & Brown, 1999; Islam & Hewstone, 1993; Stephan & Stephan, 1985). In contrast, DIM offers a more balanced approach, maintaining subgroup differentiation while promoting a superordinate identity, thereby reducing the risks of increased anxiety and promoting the generalization of attitudes beyond those directly involved in the contact experience. Compared to the Common Ingroup Identity Model (CIIM), which advocates for the recategorization of former ingroup and outgroup members under a single superordinate category, DIM avoids the identity threat and resistance often associated with the perception of forced assimilation (González & Brown, 2006). By allowing individuals to simultaneously recognize their original subgroup identity and a broader inclusive identity, DIM reduces intergroup bias without requiring the suppression of subgroup affiliations, making it a particularly effective framework for intergroup interventions. Empirical evidence supports these advantages. González and Brown (2003) demonstrated that when both subgroup and superordinate identities were made salient, participants exhibited greater generalization of positive intergroup attitudes compared to conditions where only one identity level was emphasized. This finding highlights the unique capacity of DIM to foster long-term attitude change by ensuring that positive contact effects extend beyond the specific interaction and contribute to more inclusive intergroup perceptions over time. Given these strengths, the DIM provides a more comprehensive and effective framework for reducing intergroup bias, balancing differentiation and inclusion in a way that neither decategorization nor single-category recategorization models fully achieve.

1.2. From direct to indirect contact

The effectiveness of contact interventions varies depending on social stigma, the measures considered, and the target groups. Strategies based on direct contact and empathy have shown significant results (Tassinari et al., 2022a), but they can be challenging to implement in contexts of segregation or conflict (Tassinari et al., 2022b), and a general difficulty is the feasibility of intervening in real-world contexts. Therefore, alternatives that utilize indirect contact forms, such as extended, imagined, vicarious, and parasocial contact, have been proposed (Di Bernardo et al., 2017).

More specifically, according to Wright et al. (1997), mere knowledge or observation of a positive interaction between ingroup and outgroup members is sufficient to reduce prejudice. Vezzali et al. (2014) differentiated between knowing about a cross-group interaction or observing it, labeling the first contact form as *extended* contact and the second as *vicarious* contact. *Imagined* contact is an intervention strategy that mentally simulates positive intergroup interactions. Crisp and Turner (2012) argued that simply imagining an ingroup member positively interacting with an outgroup member improves attitudes. This has many advantages over other contact forms, perhaps the most obvious being its extreme flexibility. *Parasocial* contacts predict that exposure to

outgroups in the media may be analogous and function similarly to face-to-face intergroup interactions (Park, 2012). Exposure to outgroups through media can reduce prejudice because people form parasocial relationships—emotional bonds with fictional characters or celebrities that resemble real-life friendships. When these onscreen outgroup friends feel like real-life friends, they can help reduce prejudice (Bond, 2021; Schiappa et al., 2005).

1.3. Indirect contact in the digital age: E-contact paradigm

Technological innovation and social media have enabled new forms of indirect contact in digital environments, commonly referred as online contact (White et al., 2015). In this regard, *E-contact* interventions allow structured interventions to reduce prejudice, utilizing cooperative online environments with a common goal (Gaertner & Dovidio, 2014). The term *"E-contact"* encompasses all interaction between individuals via digital platforms. More specifically, *E-contact* paradigms represent innovative methodologies in psychosocial research to study how prejudice between different groups can be mitigated through digital communication forms. These approaches focus on interactions between people from different social, ethnic, religious, or national groups, using technologies like the internet, social media, email, online chats, video calls, and other digital communication tools (White et al., 2020).

The main aim of this research line is to explore how online interactions shape people's perceptions, attitudes, and behaviors toward members of other groups. It seeks to understand how and when digital interactions can help reduce prejudice, stereotypes, and discrimination while fostering meaningful and lasting social integration.

One key benefit of *E*-contact is its accessibility, allowing intergroup contact despite geographical barriers or other challenges. Online interactions can also lower the anxiety often associated with face-to-face meetings, making participation more comfortable (White et al., 2020). Additionally, individuals have greater control over these interactions, deciding when and how to respond. *E*-contact interventions have already been used in educational programs to connect students from different national or cultural backgrounds (Amichai-Hamburger et al., 2015;White & Abu-Rayya, 2012; White et al., 2020), in social initiatives aimed at promoting dialogue between opposing groups (White et al., 2015), and in workplaces to improve collaboration and cohesion in international or culturally diverse teams (White et al., 2020). Empirical research has shown that, when well-structured, *E-contact* can be effective in reducing prejudice and fostering positive attitudes among diverse groups. However, the effectiveness of digital contact may vary depending on factors such as the quality of interactions, duration of contact, and participants' characteristics (White et al., 2015).

This type of intervention represents an emerging frontier in psychosocial research, effectively evolving Allport's (1954) hypotheses and models through new technologies (White et al., 2020). Specifically, *E*-contact interventions follow a structured, two-phase approach designed to optimize intergroup contact and ensure the generalization of positive attitudes beyond the immediate interaction. In the first phase, participants engage in self-disclosure activities, sharing personal interests, values, and experiences. This step is crucial as it allows individuals to perceive each other as unique persons rather than solely as representatives of their respective groups, reducing intergroup anxiety and fostering initial interpersonal connections (White & Abu-Rayya, 2012; White et al., 2020). However, for the contact to effectively promote intergroup bias reduction, the salience of group memberships must be introduced following this initial stage of personal acquaintance. Research suggests that when group identities are explicitly made salient early in the interaction, they remain cognitively active even if not continuously reinforced, ensuring that positive attitudes extend beyond the specific individuals involved in the exchange (Brown & Hewstone, 2005; Hewstone et al., 2006). This can be achieved through researcher-facilitated prompts, explicit group affiliation statements, or structured discussion topics that highlight group-based perspectives while maintaining the interpersonal connection established in the first stage. The second phase of the intervention involves a cooperative task in which participants must work together toward a shared goal. This phase reinforces a superordinate identity, emphasizing collaboration and commonality while simultaneously preserving subgroup distinctiveness (White et al., 2020). The deliberate activation of both subgroup and superordinate identities aligns with the theoretical principles of the Dual Identity Model (DIM; Gaertner et al., 1999; Gaertner & Dovidio, 2000; González & Brown, 2003; González & Brown, 2006), which posits that intergroup bias is most effectively reduced when individuals maintain their original group identities while also recognizing membership in a broader, inclusive category. Unlike models that promote either individualization (e.g., Decategorization) or full assimilation under a single identity (e.g., CIIM), DIM provides a balanced framework that prevents identity threat while ensuring that positive contact effects generalize beyond the immediate interaction (González & Brown, 2003, 2006). By first fostering interpersonal connections and then introducing and reinforcing both distinct and shared identities, *E*-contact interventions leverage the mechanisms proposed by DIM to maximize intergroup bias reduction and ensure long-term attitude generalization. Empirical evidence has shown that *E*-contact interventions incorporating DIM principles result in more significant and widespread reductions in prejudice than those that do not explicitly manage dual identity salience (White & Abu-Rayya, 2012; White et al., 2020). Through this structured approach, E-contact provides a scalable and effective model for fostering intergroup cohesion in digital environments, ensuring that intergroup attitudes extend beyond the specific individuals involved in the interaction.

In this regard, the DIM (Gaertner & Dovidio, 2000) offers a particularly appropriate framework for AI-mediated interventions. Unlike approaches that rely solely on decategorization or a common ingroup identity—each of which can obscure meaningful group distinctions or reduce the salience needed for generalization—the DIM encourages simultaneous recognition of subgroup membership and superordinate belonging. This balance is critical when the contact partner is a non-human agent, as mutual individuation is limited, and identity cues must be carefully scripted. The structured, sequenced nature of AI-driven interactions aligns well with DIM's theoretical requirements: personal disclosure can foreground the individual, while visual and verbal cues reinforce subgroup membership, and a shared task activates the superordinate identity. This integration of identity layers makes DIM especially suitable for guiding structured, scalable interventions using digital humans.

1.4. AI-Driven digital humans in E-contact: Intergroup interventions through parasocial interaction

The contact model (Allport, 1954) and research adopting *E*-contact paradigms have shown that this approach effectively reduces prejudice. However, a limitation of such interventions is the need for multiple actors to be present during the interaction. To overcome this limitation, we hypothesize that integrating new E-contact interventions can be significantly enhanced by implementing "artificial humans" based on artificial intelligence (AI). Artificial humans, advanced virtual entities designed to emulate human-like behavior, appearance, and communication, have become a transformative force in human-computer interaction. These entities leverage cutting-edge artificial intelligence (AI), including generative models, to process and respond to user inputs with contextual appropriateness and a natural flow, aligning closely with established psychological theories of human communication and social interaction (Bartneck et al., 2009). AI systems underpinning artificial humans are grounded in machine learning algorithms capable of recognizing patterns, processing natural language, and generating human-like responses. These systems are often trained on vast datasets, encompassing linguistic structures, cultural nuances, and emotional expressions, which enable them to simulate human communication with increasing sophistication (Von der Pütten et al., 2010).

A central feature of AI-driven artificial humans is natural language processing (NLP), which enables these systems to interpret users' inputs in ways that mimic human conversation. This allows artificial humans to engage in meaningful dialogues, respond to nuanced questions, and maintain conversational coherence across diverse topics (Nass & Moon, 2000).

Artificial humans represent a significant advancement over traditional text-based chatbots, an early and relatively simple form of conversational agents. While early text-based chatbots have demonstrated utility across various domains such as for therapeutic settings (Bickmore & Picard, 2005) or as interactive tutors (Johnson et al., 2000), their capabilities remain limited compared to the sophisticated potential of AI-powered chatbots and digital humans.

Recent research (Crabtree et al., 2024) has explored the potential of AI-based chatbots to reduce prejudice and promote support for marginalized groups, particularly transgender individuals. Using a GPT-4-based chatbot, the authors framed personalized persuasive arguments grounded in moral values, such as highlighting justice and equal treatment or emphasizing the harm caused by discrimination. Participants were randomly assigned to either a treatment group, which interacted with the chatbot, or a control group with no intervention. The study found significant short-term effects, with participants in the treatment group displaying increased support for transgender rights immediately after the intervention, including higher levels of policy support and general warmth toward transgender individuals. Although the study did not find long-term effects, these findings provide promising evidence that AI-driven, personalized persuasion could be a scalable tool for reducing prejudice, delivering meaningful short-term results comparable to traditional methods.

Despite the novelty of the findings, this work still relies on chatbots that, although powered by generative AI, continue to feature a text-based interface. In this regard, previous literature suggests that text-based chatbots are often evaluated based on anthropomorphic qualities, such as language style, responsiveness, and conversational behavior. While these human-like attributes can enhance user engagement and foster emotional connections, they can also present challenges. Chatbots that "pretend" to be human risk unintentionally deceiving users, potentially leading to trust issues or inflated expectations of their capabilities (Rapp et al., 2021). Artificial humans, by contrast, have the potential to overcome these limitations by offering more natural and socially engaging interactions. Equipped with advanced AI, artificial humans can facilitate more realistic conversations and, unlike traditional chatbots, they can potentially eliminate the need for a human counterpart in contact interventions, offering scalable and consistent solutions for intergroup interactions. Digital humans can be customized to fit diverse cultural and linguistic contexts, ensuring interactions are relevant, accessible, and meaningful for different populations. Their scalability, combined with cost-effectiveness and efficiency, makes them a powerful tool for fostering intergroup contact. Like other forms of *E*-contact, they overcome traditional methods' geographical and temporal barriers, but with the added benefit of an anthropomorphic design that should enhance engagement and trust.

Indeed, while previous interventions have relied on text-based chatbots that mimic human-like communication but lack genuine agency (e.g., Crabtree et al., 2024; Sahab et al., 2024), the adoption of AI-based digital humans may help overcome this limitation by providing a more socially engaging experience. Unlike traditional *E*-contact interventions, which involve real human counterparts, AI-driven digital humans introduce a one-sided, asymmetrical interaction that aligns with parasocial contact theory (Park, 2012) Recent typologies of indirect contact place this kind of procedure within the "hybrid E-contact/parasocial" category (White et al., 2020), as it merges the reduced threat of asynchronous digital exchange with the immediacy and relevance of direct address. This framing helps situate the intervention described below within a broader theoretical effort to understand how digital interfaces can evoke meaningful intergroup engagement through varying combinations of disclosure, responsiveness, and perceived social connection. This theoretical framing also highlights why parasocial closeness, rather than constructs tied to mutual interaction like empathy or perceived similarity, offers a more fitting lens for interpreting the effects of AI-mediated contact.

Parasocial closeness describes the phenomenon where individuals develop a sense of connection and familiarity with media figures - or in this case artificial agents - despite the absence of reciprocal exchange (Park, 2012). Although these interactions lack social contingency, they can still evoke emotional closeness and a sense of social presence (Bond, 2021; Schiappa et al., 2005). In the context of AI-driven contact interventions, the degree of parasocial closeness participants experience with the digital human is expected to be a key psychological mechanism influencing changes in intergroup attitudes. In this regard, prior research has demonstrated that parasocial bonds can facilitate prejudice reduction by lowering intergroup anxiety, enhancing outgroup perspective-taking, and fostering greater openness to attitude change (Schiappa et al., 2005; Vezzali et al., 2015). Specifically, individuals who perceive AI-driven digital humans as socially meaningful partners may experience greater empathy and reduced perceived social distance, ultimately shaping their real-world intergroup perceptions (Bond, 2021; Chen & Zhang, 2022; Schiappa et al., 2005). This mechanism becomes particularly relevant when the digital human represents an outgroup member, as it can encourage participants to engage in more personalized and emotionally resonant interactions, thereby reinforcing the intervention's effectiveness.

The investigation of parasocial closeness as a possible intervening variable could, therefore, extend prior research on AI-driven social interactions, offering new insights into the psychological mechanisms that drive the effectiveness of artificial humans in intergroup contact interventions.

In this context, parasocial closeness captures a distinctive mechanism of relational engagement with artificial partners. Unlike empathy, which refers to a transient emotional alignment with another's internal state, or social presence, which reflects the immediate perception of a responsive other (Nowak & Biocca, 2003), parasocial closeness entails a perceived, lasting social bond with a non-human partner. It involves the feeling of "knowing" and trusting the digital agent across interactions, despite the awareness of its artificiality (Bond, 2021; Schiappa et al., 2005). This makes parasocial closeness especially relevant for AI-mediated intergroup interventions, where users may engage meaningfully with a partner they recognize as non-human but still perceive as socially significant. Moreover, it helps conceptually locate our intervention between imagined contact—entirely mental and unidirectional—and extended contact, which lacks direct interaction: unlike imagined contact, participants in our paradigm see, hear, and respond to a lifelike agent; unlike extended contact, they form a personalized connection. For these reasons, in the study described below, parasocial closeness was selected as a mechanism underlying attitude change, while still monitoring empathy and social presence descriptively to capture adjacent experiential dimensions.

2. Overview

The study aims to assess the impact of an *E*-contact intervention on members of the host society by involving them in a three days-long interaction with an AI-based artificial human representing an outgroup or ingroup member. Grounded in Allport's (1954) conditions for prejudice reduction and White's E-contact model (White et al., 2020), the intervention includes three scripted activities (Maunder et al., 2019) designed to foster personal acquaintance (Brewer & Miller, 1984) and common group affiliation (Gaertner et al., 1989, 1993; Gaertner & Dovidio, 2014), ensuring that participants experience both subgroup differentiation and superordinate categorization. This special *E*-contact intervention, like the one proposed and described by White and Abu-Rayya (2012), is theoretically based on the Dual Identity Model (DIM)

(Gaertner et al., 1999; Gaertner & Dovidio, 2000; González & Brown, 2003, 2006).

Furthermore, the intervention is structured to meet Allport's optimal contact conditions as described below. As outlined above, to ensure equal status, the virtual assistant is portrayed as a university student, similar to the participants. Cooperation toward a shared goal is promoted through joint activities, with institutional support explicitly conveyed during the interaction. We hypothesize that engaging in structured conversations with an AI-driven artificial human representing an outgroup member will lead to a significant improvement in both explicit and implicit attitudes toward the target outgroup, consistent with White's (White et al., 2020) *E*-contact framework.

More specifically, participants in the outgroup condition are expected to show greater reductions in intergroup bias compared to participants in the ingroup condition. While the main hypothesis concerns the direct impact of avatar condition on prejudice reduction, parasocial closeness will be also preliminary explored as a potential moderator of this effect.

3. Method

3.1. Ethics & inclusion statement

The present study has been carefully designed to minimize any risk of stigmatization, incrimination, discrimination, or personal harm to participants. Provisions have been implemented to ensure the safety and well-being of all individuals involved, in full compliance with ethical standards. The research protocol has been approved by the Ethics Committee for the evaluation of minimal-risk studies at the Department of Psychology of the affiliated university (Ref. #RM-2024-896).

Prior to participation, individuals will receive comprehensive information about the study and their rights. The informed consent process will emphasize voluntary participation, the right to withdraw at any time without penalty, and the assurance of confidentiality. No sensitive personal data will be collected; only non-identifiable interaction data will be recorded, in full accordance with institutional and data protection policies. All data will be securely stored and anonymized to protect participants' privacy.

At the end of the study, participants will be thoroughly debriefed and informed about the broader aim of the research—namely, the promotion of positive intergroup interactions across different ethnic groups. Additionally, the web-based platform used in the study, developed by QuestIT (Algho), has been made available free of charge by the company. The platform does not store any sensitive personal data.

3.2. Participants and procedure

To examine the effects of interaction with a digital human on prejudice reduction, a between-within mixed-design experimental approach will be adopted. The sample size was determined using G*Power analysis software, based on the effect size reported in the meta-analysis by Imperato and Mancini (2022; $f = 0.18$; $d = 0.36$), to determine the necessary sample size for a 2 (white vs black partner) × 2 (pre and post-intervention) mixed-design ANOVA (within-between interaction), considering $\alpha = 0.05$, power $(1 - \beta) = 0.80$, and a correlation among repeated measures of 0.50. The results indicated that a total sample size of 64 participants (i.e., 32 per group) would be required to achieve adequate statistical power. Although no gender effects are expected, the sample will be balanced by gender. To account for potential dropouts or technical issues, the study will include 70 Caucasian university students. In this regard, it is worth noticing that the present study focuses on majority-group members because majority attitudes often drive institutional gate-keeping and because prior DIM work identifies majority resistance to subgroup salience as a key barrier to change (Dovidio et al., 2007). This choice delivers a stringent test of the dual-identity script.

Participants will be recruited through the Department of Psychology's recruitment platform (i.e., Sona System) and will receive academic credits in exchange for their participation. To ensure consistency in avatar assignment and to avoid the risk of misalignment between participant identity and avatar representation, only individuals who self-identify as male or female will be eligible to participate in the current study. This restriction is necessary because the Algho © platform presently offers only binary-gender avatar templates (male and female), and does not support customization of body morphology or the creation of gender-neutral digital humans. Potential participants will be clearly informed of this eligibility criterion during the recruitment process. As a cover story, they will be informed that the study involves testing the features of a new AI agent-based application. They will be told that the aim is to evaluate various aspects of human interaction with artificial intelligence agents. To emphasize institutional support, participants will also be told that the study is part of a broader research project on assistive technologies and social inclusion, funded by the European Union.

Participants will first be invited to the laboratory to complete a preliminary questionnaire, which will collect demographic information and assess key psychological constructs prior to the experimental manipulation. Following this, they will be randomly assigned to one of two experimental conditions. Over the course of three days, participants will be asked to engage daily—using their smartphones—with either a Black or White AI-based digital human. In one condition (Black avatar), the digital human will report being Nigerian, while in the control condition (White avatar), the digital human will report being from another city in the participant's country of origin (see also the Supplementary material). After three days, participants will be invited back to the lab for assessing the post-manipulation questionnaire.

These digital humans will be AI-controlled 3D avatars, with the avatar's gender matched to that of the participant to ensure consistency in gender representation and reduce potential confounding effects (see Fig. 1). The four artificial humans (Black male/female and White male/female) have been developed by the company QuestIT using Algho, a web-based no-code platform specifically designed for creating 3D AI-driven digital assistants. Algho enables the development of avatars with advanced features that support interactive and empathetic virtual communication (see Fig. 1). Since the adopted digital humans are drawn from a pre-configured library provided by the Algho platform, their appearance cannot be altered beyond the parameters required for the group manipulation (i.e., skin tone and hair texture). To mitigate the risk that residual visual differences between avatars might influence participants' responses, an additional control step will be implemented: participants will be asked to evaluate each avatar on a set of key social traits following the interaction. These ratings will be examined and statistically controlled in the main analyses if needed, to ensure that the interpretation of experimental effects is not confounded by unintended differences in avatar perception.

To ensure consistent participation, a smartphone application specifically developed for conducting longitudinal study (i.e., Time2Rate)–will be used. Through daily notifications sent via the app, participants will be reminded to interact with the virtual assistant once a day for the three days of the manipulation period, with direct access provided through a link in the notification. Notifications will be sent every morning throughout the study period and consist of a message saying, "[NAME OF THE ARTIFICIAL HUMAN] has sent you a message, chat with him/her," followed by an access link for direct interaction with the AI-based human. To monitor consistent and effective participation without compromising their privacy, participants will be assigned a unique numerical identification code, which they will be required to enter at the beginning of each activity. Additionally, as a manipulation check, the app will allow tracking of the actual frequency of interactions with the virtual assistant.

The digital human is designed to follow a specific script tailored to each day of interaction. In practice, this means the conversations initiated by the digital human are guided and follow a semi-structured framework. A dedicated AI module for natural language processing

Fig. 1. Screenshots of the AI-based artificial humans with whom participants will interact for three days, based on gender identity. The AI-based virtual assistant was designed with realistic facial features, culturally appropriate attire, and neutral expressions to enhance reliability and authenticity. The avatar's gender was matched with participants to ensure consistency and minimize potential gender-based biases.

enables the system to interpret participants' responses and smoothly redirect the conversation toward the key points of the script. This approach ensures consistency across participant interactions while allowing the AI to adapt to user inputs and keep the dialogue aligned with the intended narrative of the experimental design. Additionally, this approach effectively handles any deviations participants may introduce, redirecting the dialogue to the established and standardized experimental framework. Moreover, to ensure cultural sensitivity, the AI-based digital human was designed with input from experts in intercultural communication. The avatar's language and behavior were informed by validated frameworks to avoid reinforcing stereotypes while fostering authentic and respectful representation. Specifically, we relied on ethical AI guidelines for conversational agents (Piñeiro-Martín et al., 2023; Ruane et al., 2019) to ensure that the artificial humans used neutral, non-stereotypical language across all conditions. These frameworks emphasize the importance of avoiding variations in politeness, assertiveness, or response complexity based on social categories like gender or ethnicity.

All interactions will occur via typing on the smartphone keyboard in a chat format, while the artificial human will respond and interact using speech synthesis. This approach was chosen to prevent connectivity issues and potential interaction difficulties for participants.

The intervention will span three consecutive days of interaction with an artificial human, following a structured two-phase approach designed to optimize intergroup contact in line with the *E*-contact paradigm (White & Abu-Rayya, 2012). In each session, participants will first engage in a self-disclosure conversation, starting with personal experiences and gradually introducing the salience of group memberships. In the second phase, a superordinate group identity will be activated, followed by a cooperative task in which participants and the artificial human collaborate to achieve a shared goal. While each interaction will follow a scripted format to ensure standardization and alignment with the *E*-contact model, the scripts will differ across the three days. Specifically, although the overall structure remains constant, the content of the cooperative activities will vary from day to day.

To select the most effective scripts, a pilot study will be conducted in which five different scripts will be evaluated based on perceived cooperation, equal status, institutional support, and common goals. Additionally, the scripts will be assessed based on the levels of categorization they activate, ensuring alignment with the *E*-contact paradigm (White & Abu-Rayya, 2012; White et al., 2020). Participants will read and assess all five scripts in a randomized order to control for potential order and anchoring effects. To determine the adequacy of the sample size for the pilot study, an a priori power analysis using G*Power 3.1 (Faul et al., 2007) was conducted. The analysis was based on a within-subjects design with a five-level factor, corresponding to the repeated evaluation of five experimental conditions. The calculation was performed for a one-factor repeated-measures ANOVA, assuming a correlation between repeated measures of $\rho = 0.50$ and a significance level of $\alpha = 0.05$. The expected effect size was set at $f = 0.25$, corresponding to a *partial* η^2 of approximately 0.06, representing a small effect size according to Cohen's criteria (1988). To achieve a statistical power of 0.80, the analysis indicated that 21 participants would be required. Therefore, a total of 25 individuals will be invited to participate in the pilot test, taking potential dropouts into account.

Following this pre-study, the three scripts receiving the highest and most equivalent scores across the target dimensions will be selected for the main intervention, ensuring optimal engagement and effectiveness.

3.3. Measures

The following constructs will be assessed in the preliminary survey:

3.3.1. Demographic measures

To provide a comprehensive overview of the sample, participants will complete a series of demographic questions. These questions will assess key variables such as age, gender, occupation, and education level.

3.3.2. Political orientation

Participants' political orientation will be assessed using a sliding scale ranging from 1 to 101, where 1 represents “very left-wing” and 101 represents “very right-wing.” (Kroh, 2007).

3.3.3. Familiarity with AI and AI-based digital assistants

Familiarity with AI tools will be assessed as a control variable by adopting three ad-hoc created items already used in previous research (see Gabbiadini et al., 2024). The items will be as follows: 1) “How

would you rate your level of experience/familiarity with AI technologies?" anchored with 1 = *no experience*, 2 = *beginner*, 3 = *competent*, 4 = *expert*, 5 = *very experienced*, 2) "How would you rate your knowledge of AI technologies?" anchored with 1 = *no knowledge*, 2 = *limited*, 3 = *average*, 4 = *good*, 5 = *very good* and 3) "Have you ever used AI based assistants?" anchored with 1 = *never*, 2 = *once*, 3 = *occasionally*, 4 = *several times*, 5 = *very often*.

The following constructs will be assessed in both the pre and post manipulation surveys.

3.3.4. Implicit attitudes toward black people

Implicit attitudes will be assessed using a standard IAT with seven blocks that measure association strengths between racial categories (Whites and Blacks) and pleasant and unpleasant words (Xu et al., 2014), a well-established measure of implicit intergroup bias, demonstrating robust reliability and predictive validity in assessing automatic racial attitudes across diverse samples (Greenwald et al., 2009; Kurdi et al., 2019). The White and Black stimuli will be face images. Specifically, the seven blocks will include two 20-trial single categorization practice blocks (Blacks vs. Whites; positive vs. negative words), followed by a 20-trial and then a 40-trial combined categorization block (e.g., Black and positive words vs. White and negative words), then another single categorization block with 40 trials and category sides switched (White vs. Black), and finally another two combined-task blocks of 20 and 40 trials (White and positive words vs. Black and negative words).

3.3.5. Explicit attitudes toward black people

Participants will complete semantic differential scales for Blacks, semantic differential scales for Whites, a feeling thermometer for Blacks, and a feeling thermometer for Whites, in that order (McConnell & Leibold, 2001). Seven-point scales will be used for the semantic differential word pairings: beautiful–ugly, good–bad, pleasant–unpleasant, honest–dishonest, and nice–awful.

Participants will also report their attitudes toward Blacks and Whites using a feeling thermometer, which will range from 0° (*extremely unfavorable*) to 100° (*extremely favorable*). The intergroup bias index will be calculated for the different attitude measures employed.

After one week, explicit and implicit attitudes toward the outgroup will be measured once again, using the same instruments as those described above. The following constructs will be only assessed in the post manipulation survey.

3.3.6. Manipulation check

As a manipulation check, two items will be used to understand how participants categorized themselves and the artificial human. A self-categorization measure adapted from Gaertner et al. (1989) will be used to assess the extent to which participants perceived both subgroup distinctiveness and shared superordinate identity, in line with the principles of the Dual Identity Model (DIM). They will be asked to select the representation (one group, two distinct groups, separate individuals, one group composed by two subgroups) that best reflects their perception of the aggregate during the interaction ("During the interaction, did you perceive yourself and the avatar in the environment as members of one group, as members of two different groups, as separate individuals, or as members of a social group composed by two subgroups?") and to what extent on a 7-point scale. Participants' categorization of the black avatar will be assessed using an ad hoc created single item: "To what extent do you perceive the avatar you interacted with as belonging to your group or to a different group?" Responses will be measured on a 7-point scale (1 = *belonging to my social group*; 7 = *belonging to a different social group*). This categorical measure of identity endorsement will be used as a manipulation check to assess whether participants perceived the intended dual identity activation during the avatar interaction. Specifically, the proportion of participants who selected each response option (e.g., "two subgroups within one group," "two distinct groups") will be reported. This check allows to verify that the contact sequence is aligned with the Dual Identity Model (DIM) framework. However, identity endorsement will not serve as a mediator or as a formal condition for hypothesis confirmation, as the main focus remains on testing whether the avatar condition (ingroup vs. outgroup) leads to differences in prejudice reduction.

Moreover, to assess whether participants perceive Allport's (1954) optimal contact conditions 8 items will be included at the end of the interaction period. Participants will respond to each of the following statements on a 7-point Likert scale (1 = *strongly disagree*, 7 = *strongly agree*):

Perception of Equal Status: *"I felt that the virtual assistant and I were on the same level during our conversations."* And *"The virtual assistant and I had an equal role in the interaction."*. Perceived levels of Cooperation: *"I felt like the virtual assistant and I were working together as a team."* And *"The activities with the virtual assistant required us to collaborate."*. Common Goals: *"The virtual assistant and I were working toward the same goal."* And *"I felt like we had a shared purpose during the activities."* Perception of Institutional Support: *"My interaction with the virtual assistant was encouraged and supported by the University and the European Union"* and *"I felt that the virtual assistant was introduced as part of a structured and well-organized intervention by the University and the European Union."*

3.3.7. Retrospective parasocial closeness

Two factors from the scale developed by Tukachinsky (2010) will be utilized to measure parasocial relationships. Specifically, the factors related to friendship-like relationships will be assessed: the Friendship Communication factor, which includes dimensions such as disclosure and advice-seeking behaviors, and the Friendship Support factor, which captures perceived emotional and social support. Sample items are "If the artificial human was a real person, I could have disclosed negative things about myself honestly and fully (deeply) to him/her" and "If the artificial human were a real person, I would give him/her my emotional support" (all measured on a 7-point scale from 1 = *completely disagree* to 7 = *completely agree*). These factors were chosen because they reflect key components of parasocial interactions, resembling real-world friendship dynamics and providing a nuanced understanding of participants' connections with the AI-based digital human.

3.3.8. Perceived naturalness of the interaction

Prior research (So et al., 2023) suggests that repeated interactions with AI can influence perceptions of realism and credibility, which may, in turn, affect the dependent variables. To account for this, a set of items has been developed to assess how participants perceive the naturalness of their interaction with the AI-based digital human. These items are designed to measure realism, human-likeness, and conversational credibility, key factors in determining whether the digital human is experienced as a genuine social partner or merely as a scripted artificial agent.

Building on previous literature (So et al., 2023), the following ad hoc items will be assessed on a 7-point Likert scale ranging from 1 = Completely Disagree to 7 = Completely Agree. To measure perceived realism, participants will respond to the items: *"The virtual assistant's responses felt natural and spontaneous."* and *"The interaction with the virtual assistant resembled a conversation with a real person."* For perception of human-likeness, the following items will be used: *"The virtual assistant displayed behaviors and expressions similar to those of a human."* and *"I sometimes forgot I was interacting with an AI rather than a real person."* Finally, to assess perceived conversational credibility, participants will evaluate the items: *"The virtual assistant's responses were coherent and made sense in the context of our conversation."* and *"I found the virtual assistant's reactions to be appropriate and relevant to what I said."*

3.3.9. Post-interaction perceived social traits

To account for the possibility that subtle visual cues in the avatars may inadvertently influence participants' impressions, we will assess perceived social traits of the digital avatars after the VR interaction. This

measure is designed to detect differences in perceived warmth, trustworthiness, competence, and likability, which could serve as potential confounds if they systematically vary across avatars. Participants will rate the avatar they interacted with on a series of 7-point semantic differential scales, adapted from previous work on anthropomorphic facial trustworthiness and perceived social appeal (Brownlow, 1992; Song et al., 2021). The items include *Warm – Cold, Trustworthy – Untrustworthy, Competent – Incompetent* and *Likeable – Unlikeable.* These ratings will be administered at the end of the post-interaction questionnaire. If significant differences emerge in trait evaluations across avatars, these variables will be statistically controlled in the main analyses to ensure that primary results are not confounded by avatar-specific appearance effects.

3.3.10. Controlling measures of empathy and social presence

To complement the primary analyses, we will include a short set of items to descriptively assess participants' empathy and social presence during the interaction. These constructs will not be included in the main mediation model but will serve to contextualize participants' subjective experiences and explore their relationship with parasocial closeness. A subset of items from the State Empathy Scale (Shen, 2010) will be adapted to reflect participants' affective and cognitive engagement with the avatar while avoiding anthropomorphic assumptions. Sample items include: "*I tried to put myself in the digital human's shoes*" and "*I imagined what it might be like to experience what the digital human described.*" In addition, to assess social presence, we will adapt items from the Networked Minds Social Presence Scale (Biocca et al., 2001), focusing on mutual awareness and involvement within a non-immersive, smartphone-based context. Sample items include: "*I felt that the digital human acknowledged my input*" and "*The digital human made me feel engaged in the exchange.*" All items will be rated on a 7-point Likert scale (1 = not at all, 7 = very much).

3.4. Data analysis strategy

The primary hypothesis concerns the Condition × Time interaction predicting prejudice reduction across three dependent variables: the IAT D-score, semantic-differential bias, and the feeling thermometer. For each outcome, a 2 (condition: ingroup vs. outgroup avatar) × 2 (time: pre vs. post) mixed ANOVA will be conducted. To account for multiple comparisons, a Bonferroni correction will be applied, adjusting the alpha level to $p \leq 0.0125$ for the interaction term in each ANOVA. Parasocial closeness will be examined as an exploratory moderator of change. This analysis will be conducted only if a significant Condition × Time interaction emerges in the Black-avatar group. The MEMORE macro (Montoya, 2019; Model 1 with 5000 bootstrap samples) will be used to test moderation, with parasocial closeness entered as a continuous moderator of the pre–post change score. This macro allows for estimating total, direct, and indirect effects while accounting for the dependency of repeated measures data. This analytical framework will enable us to evaluate whether the pre-to-post change in attitudes operates through the proposed intervening variable while controlling for potential confounds.

To address well-documented concerns regarding the temporal stability and retest sensitivity of the IAT in within-subject designs (e.g., Nosek et al., 2007), we will compute test–retest reliability by correlating pre- and post-intervention D-scores. This check will help determine whether observed changes reflect genuine effects or random measurement fluctuation. We will also implement a set of robustness checks to ensure the reliability of results. Specifically, we will exclude extreme IAT D-scores (±2.5 SD from the sample mean).

Additionally, post-interaction ratings of the avatar's warmth, competence, and authenticity will be inspected. If these traits vary systematically by condition, they will be included as covariates and tested for interaction effects with the avatar group, following Dovidio et al. (1998). Finally, political orientation and familiarity with AI will be included as covariates in the repeated measures ANOVA to account for potential confounding effects. These variables will also be analyzed through correlation analyses to explore their independent relationships with changes in intergroup attitudes. Empathy and social presence will be reported in an online appendix and explored in supplementary tests only if sample size allows.

Missing data are not expected, as all measures are completed in-lab during the same session. If minor omissions occur, they will be handled via list-wise deletion, and the final *N* will be transparently reported for each analysis.

3.5. Exclusion criteria

Since the current research aims to examine intergroup interactions from the perspective of majority group members, only participants of Caucasian ethnicity will be involved in the study. This methodological decision is necessary to maintain the study's integrity and focus. The rationale for participant selection will be clearly communicated, to ensure transparency and in line with the ethical principles outlined in the National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research (1979) and the APA Ethics Code (2017), which emphasize that the inclusion of participants should be based on scientific necessity rather than arbitrary exclusion. Additionally, participants will be excluded from the final analysis if they fail to complete all required activities within the intervention or if they leave any tasks unfinished. Furthermore, individuals who do not adhere to the structured conversation framework when interacting with the artificial human will be excluded, as deviations from the expected interaction patterns could compromise the validity of the experimental manipulation. Another exclusion criterion concerns the manipulation checks assessing participants' perception of Allport's (1954) optimal contact conditions, specifically equal status, cooperation, common goals, and institutional support. Since the *E*-contact paradigm is structured to ensure that these conditions are met, it is essential to verify that participants have perceived the interaction accordingly. Failure to do so would indicate a misalignment between the intended structure of the intervention and participants' subjective experience, potentially compromising the validity of the findings.

3.6. Expected results

We expect that interacting with an AI-based 3D artificial human representing a Black individual, compared to a White AI-driven digital human, will reduce intergroup bias. This aligns with E-contact research, which has shown that mediated, structured intergroup interactions can foster positive attitudes by reducing social distance and intergroup anxiety (Vezzali et al., 2015; White & Abu-Rayya, 2012; White et al., 2020). Although most E-contact studies focus on human-human interactions, emerging evidence suggests that AI-driven conversational agents may similarly influence intergroup attitudes by reducing inter-ethnic anxiety and improving outgroup perceptions (Sahab et al., 2024). Additionally, parasocial relationships with AI-driven entities may facilitate attitude change in ways comparable to human-mediated parasocial contact (Dibble et al., 2016; Wong et al., 2022). Given these premises, we anticipate changes in both implicit and explicit intergroup attitudes, highlighting the potential of AI-based *E*-contact for prejudice reduction. Since prior research suggests that parasocial relationships with outgroup media figures can foster emotional connection and perspective-taking, contributing to prejudice reduction (Bond, 2021; Schiappa et al., 2005), we expect that the anticipated reductions in intergroup bias will be moderated by relationship closeness with the AI-driven digital human and the development of parasocial relationships. We consider the possibility of alternative outcomes, where the effectiveness of AI-driven contact may vary based on individual differences or the need for further intervention optimization. Participants with prior AI experience or a higher receptiveness to digital interventions may

demonstrate more pronounced or rapid changes in attitudes (Sahab et al., 2024). Additionally, the extent of intergroup attitude change may depend on participants' baseline attitudes, with initial biases or predispositions potentially influencing the intervention's effectiveness. Conversely, if no significant effects will be observed, this could indicate that factors such as intervention duration, personalization, or conversational depth require further refinement to enhance the impact of AI-driven contact. Alternatively, null effects might reflect participants-related factors, such as novelty effects — where initial participant engagement is driven by the uniqueness of interacting with an AI partner rather than by substantive intergroup processes — or technological skepticism, whereby participants' distrust or discomfort with AI-mediated interactions diminishes the intervention's effectiveness.

4. Conclusion

This study represents an initial exploratory step in advancing *E*-contact research by introducing an AI-driven intervention that utilizes generative AI-powered digital humans to facilitate structured, immersive intergroup interactions. While traditional *E*-contact studies have primarily focused on human-human interactions, this study begins to extend the theoretical boundaries of the field by examining whether AI-driven outgroup avatars can elicit similar psychological mechanisms of attitude change, offering preliminary insights into the flexibility of intergroup contact theory in digitally mediated environments. Beyond its theoretical contributions, this study highlights the potential real-world applicability of AI-driven *E*-contact interventions. Their scalability and adaptability suggest that, if effective, they could provide a promising complement to human-facilitated interactions, offering structured and customizable engagement across multiple domains. In the educational sector, AI-driven E-contact interventions could, with further validation, be integrated into diversity and inclusion programs in schools and universities, allowing students to engage in structured intergroup dialogues within a controlled, psychologically safe environment. These interactions could also be embedded in digital learning platforms, where students participate in interactive exercises designed to challenge biases and promote perspective-taking. In professional environments, AI-based intergroup interactions might be incorporated into onboarding processes, continuous professional development, and organizational change initiatives to foster inclusive workplace cultures. Simulated intergroup interactions could potentially help employees navigate diverse work environments and improve intergroup dynamics. From a social integration perspective, AI-driven interventions could eventually support migrant and refugee inclusion programs by providing accessible, structured engagement to facilitate language practice, cultural adaptation, and social connection. However, the scalability and generalizability of AI-driven *E*-contact interventions remain to be established through systematic empirical testing.

To evaluate whether this form of AI-mediated contact yields meaningful change, we will interpret observed effects in light of benchmarks established in recent meta-analytic literature on digital intergroup contact. Pereira da Costa and collegues (2024), reviewing studies involving computer-mediated, asynchronous, and immersive contact formats (e.g., text-based chats, virtual environments), report an average effect size of $g = 0.25$ across both implicit and explicit measures of prejudice. Imperato et al. (2021), focusing on *E*-contact and other online modalities, estimate a pooled effect size of $d = 0.36$. Based on these reference points, any observed effects within or above this range would constitute a positive outcome, providing initial support for the feasibility and theoretical relevance of AI-based interventions grounded in the Dual Identity Model. Such findings would suggest that AI-driven digital humans can activate psychological processes of bias reduction at levels comparable to those achieved by more established digital contact paradigms. If both avatar conditions will produce equivalent effects, despite participants having categorized the Black avatar as an outgroup member, this outcome would point to possible methodological limitations in the implementation of the group membership manipulation. One possible interpretation is that, although participants formally recognized the avatar's group membership, the social category may not have been sufficiently salient, emotionally engaging, or consequential within the structure of the interaction to elicit differentiated attitudinal change. Previous research indicates that the salience and psychological relevance of group membership are crucial preconditions for contact interventions to selectively reduce outgroup bias (Pettigrew & Tropp, 2006). Future studies should therefore focus on strengthening the emotional resonance and social meaningfulness of group boundaries during AI-mediated interactions to ensure that intergroup processes, rather than generalized interpersonal dynamics, drive attitude change. Similarly, the moderation analysis involving parasocial closeness should be interpreted with caution, as both the moderator and outcome measures are collected concurrently, limiting the extent to which causal conclusions can be drawn. In conclusion, by demonstrating the potential feasibility of structured interactions between groups based on artificial intelligence, this study aims to pave the way for the development of advanced technological solutions for reducing prejudice and promoting social cohesion in an increasingly digital and globalized world.

5. Limits and future directions

A first limitation of the present study concerns the exclusive inclusion of White participants. While this choice was theoretically motivated by the goal of targeting majority-group bias dynamics, it restricts the scope of the findings to one side of the intergroup relationship. In multicultural societies, where intergroup interactions involve mutual perceptions and agency from both majority and minority group members, future research should examine how the same AI script functions for minority participants—who typically value subgroup recognition more than majority peers (González & Brown, 2006). Dual-identity contact can reduce bias for low-status actors yet increase it for high-status actors when power concerns are salient (Dovidio et al., 1998). A planned follow-up will therefore cross participant status (majority vs. minority) with avatar group to clarify whether the psychological mechanisms observed here generalize across power positions. Such efforts would enhance both the generalizability and the ecological validity of AI-driven contact interventions. Another limitation concerns the gender inclusivity of avatar representation. Due to technical constraints associated with the Algho © platform, which currently supports only binary-gender avatar templates (male and female), the present study restricted participation to individuals who self-identify as male or female. While this decision aimed to prevent identity misalignment and uncontrolled visual variance, it also excludes non-binary participants, thereby limiting the inclusiveness and generalizability of the findings. Although avatar gender was matched to the participants' self-identified gender to maintain alignment, future developments should prioritize the availability of gender-neutral avatars in order to ensure broader inclusivity and to allow participants of all gender identities to engage meaningfully in interactions involving avatar representation.

Furthermore, a key consideration in AI-guided *E*-contact interventions, and a potential limitation of this study, is the duration of the interaction. While some E-contact interventions have been implemented as single-session experiences (White et al., 2020), longitudinal studies indicate that extended and repeated interactions can reinforce intergroup bonding and produce more sustained attitudinal change (White & Abu-Rayya, 2012). Specifically, research on Dual Identity-Electronic Contact (DIEC) has shown that multi-session *E*-contact interventions, conducted over several weeks, can significantly reduce intergroup bias and intergroup anxiety, with effects persisting even after six months (White & Abu-Rayya, 2012). Given that AI-based interactions may differ in terms of emotional engagement and adaptability compared to human-human E-contact, future research should investigate whether longer or repeated AI-based interactions enhance attitude change. Additionally, examining strategies such as personalization, adaptive dialogue, and

follow-up engagement could help determine how to sustain these effects over time.

Moreover, recent evidence shows that reactions to superordinate inclusion depend on motivational fit: when ethnic identification is very strong and the personal need to belong to a broader community is weak, dual-identity messages can be met with skepticism or even perceptions of unfair treatment (Borinca et al., 2025). Our undergraduate sample is relatively homogeneous and, on average, values university membership, so most participants should fall inside the receptive window identified by that research. However, future work should test the same AI script in more diverse populations where motivational mis-matches may emerge, and dual-identity activation may be less effective or potentially counterproductive. Although DIM is appropriate in multicultural settings, environments that favour assimilation may benefit more from a one-group frame (Dovidio et al., 1998). Future studies should therefore compare dual-identity and one-group scripts across cultural contexts and status configurations. Including additional moderators—such as ethnic identity centrality and need to belong—could clarify the boundary conditions under which AI-based DIM interventions yield the strongest effects.

An additional caveat concerns the absence of a no-contact control group. Without a baseline condition, the study cannot determine whether changes in intergroup attitudes are driven specifically by the content of the outgroup interaction or by more general factors shared across both avatar conditions—such as the novelty of the AI interface, task-related engagement, or social-desirability concerns. Although the inclusion of a no-contact group would have strengthened causal inference, practical constraints related to participant time and scheduling prevented its implementation. Future studies should consider incorporating either a no-contact baseline or a minimal-interaction control (e.g., object-only conversation) to isolate the specific contribution of intergroup contact per se from these broader influences. A further consideration concerns the interpretation of the dual identity categorization measure. Although this variable helps verify whether the intervention successfully activated subgroup and superordinate identity simultaneously, it is possible that prejudice reduction occurs even when participants do not explicitly endorse a dual identity. In the current study, this measure is used descriptively to confirm that the DIM-based structure was perceived as intended. Future research could explore whether the strength of identity endorsement moderates the effects of contact and examine whether stronger or weaker endorsement of dual identity predicts more substantial attitude change.

Another important limitation concerns the scripted nature of the digital human avatars used in this study. Although the avatars were designed to minimize stereotype reproduction and support positive intergroup engagement, their behavior remains constrained by pre-programmed scripts, limiting their ability to model the full complexity and adaptivity of human social interactions. This constraint may reduce the perceived authenticity and dynamic responsiveness of the interaction, potentially impacting participants' emotional engagement and attitude change. While AI-powered digital humans offer scalability and standardization, it remains unclear whether they can achieve similar bias reduction effects compared to human-human interactions. Future research should develop follow-up studies that directly compare scripted AI avatars with human interlocutors, in order to critically assess whether AI-driven *E*-contact can produce comparable levels of social influence and bias reduction.

Additionally, further optimization efforts — such as more sophisticated conversational design or multimodal engagement strategies — may be necessary to enhance the effectiveness of AI-driven interventions. Understanding whether participants perceive AI-guided interactions as socially meaningful will also be essential in evaluating their potential as a complement or alternative to traditional *E*-contact paradigms. Finally, future research should also explore and systematically compare the effectiveness of immersive versus non-immersive formats of AI-driven E-contact interventions. Such comparisons would help clarify whether the degree of immersion—such as interactions occurring in virtual reality environments versus standard screen-based settings—plays a significant role in fostering intergroup connection, empathy, and bias reduction. Understanding these differences could provide critical insights into how technological affordances shape the impact of digital intergroup contact and inform the design of more effective, scalable interventions.

CRediT authorship contribution statement

Manfredi Anna: Writing – original draft, Software, Project administration, Methodology, Formal analysis, Data curation, Conceptualization. **Puzella Giulio:** Writing – review & editing, Data curation. **David Landi:** Writing – review & editing, Software. **Iolanda Iacono:** Writing – review & editing, Software. **Jacopo Michilli:** Writing – review & editing, Software, Conceptualization. **Gabbiadini Alessandro:** Writing – review & editing, Writing – original draft, Supervision, Methodology, Data curation, Conceptualization.

Patient consent statement

Participants for all studies will provide informed written consent before study enrollment.

Permission to reproduce material from other sources

No copyrighted material will be used in the present work.

Declaration of Generative AI and AI-assisted technologies in the writing process

During the preparation of this work, the authors used ChatGPT and DeepL exclusively to refine the English language and enhance the fluidity of certain sentences in this document. After using this tool/service, the authors reviewed and edited the content as needed and take full responsibility for the content of the publication.

Funding statement

The authors received no financial support for the research, authorship, and/or publication of this article.

Declaration of competing interest

No relevant financial or non-financial competing interests to report.

Acknowledgements

The digital humans used in this study were provided free of charge by the Research and Development division of QuestIT s.r.l.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.actpsy.2025.105129.

Data availability statement

The data that support the findings of this study will be available on the Open Science Framework web platform.

References

Allport, G. W. (1954). *The nature of prejudice*. Addison-Wesley.

American Psychological Association. (2017). Ethical principles of psychologists and code of conduct. https://www.apa.org/ethics/code.

Amichai-Hamburger, Y., Hasler, B. S., & Shani-Sherman, T. (2015). Structured and unstructured intergroup contact in the digital age. *Computers in Human Behavior, 52*, 515–522.

Bartneck, C., Kulic, D., Croft, E., & Zoghbi, S. (2009). Measurement instruments for the anthropomorphism, animacy, likeability, perceived intelligence, and perceived safety of robots. *International Journal of Social Robotics, 1*(1), 71–81. https://doi.org/10.1007/s12369-008-0001-3

Bickmore, T., & Picard, R. (2005). Establishing and maintaining long-term human-computer relationships. *ACM Transactions on Computer-Human Interaction, 12*(2), 293–327. https://doi.org/10.1145/1067860.1067867

Biocca, F., Harms, C., & Gregg, J. (2001, May). The networked minds measure of social presence: Pilot test of the factor structure and concurrent validity. In *4th annual international workshop on presence, Philadelphia, PA* (pp. 1–9).

Bond, B. J. (2021). The development and influence of parasocial relationships with television characters: A longitudinal experimental test of prejudice reduction through parasocial contact. *Communication Research, 48*(4), 573–593. https://doi.org/10.1177/0093650218816220

Borinca, I., Guerra, R., & Uka, F. (2025). "Ins and outs": Ethnic identity, the need to belong, and responses to inclusion and exclusion in inclusive common ingroups. *Group Processes & Intergroup Relations, 28*(2), 324–354. https://doi.org/10.1177/13684302241267982

Brewer, M. B. (1979). In-group bias in the minimal intergroup situation: A cognitive-motivational analysis. *Psychological Bulletin, 86*(2), 307–324. https://doi.org/10.1037/0033-2909.86.2.307

Brewer, M. B., & Miller, N. (Eds.). (1984). *Groups in contact: The psychology of desegregation*. Academic Press.

Brown, R., Vivian, J., & Hewstone, M. (1999). Changing attitudes through intergroup contact: The effects of group membership salience. *European Journal of Social Psychology, 29*(5–6), 741–764. https://doi.org/10.1002/(SICI)1099-0992(199908/09)29:5/6%3C741::AID-EJSP972%3E3.0.CO;2-8

Brown, R. J., & Hewstone, M. (2005). An integrative theory of intergroup contact. In M. P. Zanna (Ed.), *Vol. 37. Advances in experimental social psychology* (pp. 255–343). Elsevier. https://doi.org/10.1016/S0065-2601(05)37005-5.

Brown, R. J., & Turner, J. C. (1981). Interpersonal and intergroup behaviour. In J. C. Turner (Ed.), *Intergroup behavior* (pp. 33–65). Basil Blackwell.

Brownlow, S. (1992). Seeing is believing: Facial appearance, credibility, and attitude change. *Journal of Nonverbal Behavior, 16*, 101–115. https://doi.org/10.1007/BF00990325

Chen, Y., & Zhang, Y. B. (2022). Parasocial relationships and reduction of intergroup prejudice against the Chinese LGBT community: The role of intergroup anxiety and direct contact. *Journal of Applied Social Psychology, 52*(4), 204–217. https://doi.org/10.1111/jasp.12867

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Lawrence Erlbaum Associates. https://doi.org/10.4324/9780203771587

Crabtree, C., Holbein, J., Bosley, M., & Sevi, S. (2024, December 12). *Can AI reduce prejudice at scale? Evaluating the effectiveness of AI-powered personalized persuasion on support for transgender rights*. Unpublished manuscript. Dartmouth College, University of Virginia, & University of Toronto https://mbosley.github.io/papers/ai-prejudice-paper.pdf.

Crisp, R. J., & Turner, R. N. (2012). The imagined contact hypothesis. In J. Olson, & M. P. Zanna (Eds.), *Advances in experimental social psychology* (pp. 125–182). Academic Press. https://doi.org/10.1016/B978-0-12-394281-4.00003-9.

Di Bernardo, G. A., Vezzali, L., Stathi, S., Cadamuro, A., & Cortesi, L. (2017). Vicarious, extended and imagined intergroup contact: A review of interventions based on indirect contact strategies applied in educational settings. *Testing, Psychometrics, Methodology in Applied Psychology, 24*, 3–21. https://doi.org/10.4473/TPM24.1.1

Dibble, J. L., Hartmann, T., & Rosaen, S. F. (2016). Parasocial interaction and parasocial relationship: Conceptual clarification and a critical assessment of measures. *Human Communication Research, 42*(1), 21–44. https://doi.org/10.1111/hcre.12063

Dovidio, J. F., Gaertner, S. L., Hodson, G., Riek, B. M., Johnson, K. M., & Houlette, Y. (2007). Recategorization and crossed categorization: The implications of group salience and representations for reducing bias. In *Multiple social categorization* (pp. 81–105). Psychology Press. https://doi.org/10.4324/9780203969229.

Dovidio, J. F., Gaertner, S. L., & Validzic, A. (1998). Intergroup bias: Status, differentiation, and a common in-group identity. *Journal of Personality and Social Psychology, 75*(1), 109–120. https://doi.org/10.1037/0022-3514.75.1.109

Dovidio, J. F., Love, A., Schellhaas, F. M., & Hewstone, M. (2017). Reducing intergroup bias through intergroup contact: Twenty years of progress and future directions. *Group Processes & Intergroup Relations, 20*(5), 606–620. https://doi.org/10.1177/1368430217712052

Faul, F., Erdfelder, E., Lang, A. G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods, 39*(2), 175–191. https://doi.org/10.3758/BF03193146

Gabbiadini, A., Ognibene, D., Baldissarri, C., & Manfredi, A. (2024). The emotional impact of generative AI: Negative emotions and perception of threat. *Behaviour & Information Technology, 44*(4), 676–693. https://doi.org/10.1080/0144929X.2024.2333933

Gaertner, S. L., & Dovidio, J. F. (2000). Reducing intergroup bias: The common ingroup identity model. In S. Oskamp, & S. Oskamp (Eds.), *Reducing prejudice and discrimination* (pp. 1–28). Psychology Press.

Gaertner, S. L., & Dovidio, J. F. (2014). *Reducing intergroup bias: The common ingroup identity model*. Psychology Press.

Gaertner, S. L., Dovidio, J. F., Anastasio, P. A., Bachman, B. A., & Rust, M. C. (1993). The common ingroup identity model: Recategorization and the reduction of intergroup bias. *European Review of Social Psychology, 4*(1), 1–26. https://doi.org/10.1080/14792779343000004

Gaertner, S. L., Dovidio, J. F., Banker, B. S., Houlette, M., Johnson, K. M., & McGlynn, E. A. (2000). Reducing intergroup conflict: From superordinate goals to decategorization, recategorization, and mutual differentiation. *Group Dynamics: Theory, Research, and Practice, 4*(1), 98–114. https://doi.org/10.1037/1089-2699.4.1.98

Gaertner, S. L., Dovidio, J. F., Rust, M. C., Nier, J. A., Banker, B. S., Ward, C. M., & Houlette, M. (1999). Reducing intergroup bias: Elements of intergroup cooperation. *Journal of Personality and Social Psychology, 76*(3), 388–402. https://doi.org/10.1037/0022-3514.76.3.388

Gaertner, S. L., Mann, J., Murrell, A., & Dovidio, J. F. (1989). Reducing intergroup bias: The benefits of recategorization. *Journal of Personality and Social Psychology, 57*(2), 239–249. https://doi.org/10.1037/0022-3514.57.2.239

González, R., & Brown, R. (2003). Generalization of positive attitude as a function of subgroup and superordinate group identifications in intergroup contact. *European Journal of Social Psychology, 33*(2), 195–214. https://doi.org/10.1002/ejsp.140

González, R., & Brown, R. (2006). Dual identities in intergroup contact: Group status and size moderate the generalization of positive attitude change. *Journal of Experimental Social Psychology, 42*(6), 753–767. https://doi.org/10.1016/j.jesp.2005.12.006

Greenland, K., & Brown, R. (1999). Categorization and intergroup anxiety in contact between British and Japanese nationals. *European Journal of Social Psychology, 29*(4), 503–521. https://doi.org/10.1002/(SICI)1099-0992(199906)29:4

Greenwald, A. G., Poehlman, T. A., Uhlmann, E. L., & Banaji, M. R. (2009). Understanding and using the implicit association test: III. Meta-analysis of predictive validity. *Journal of Personality and Social Psychology, 97*(1), 17.

Hewstone, M., & Brown, R. (1986). Contact is not enough: An intergroup perspective on the contact hypothesis. In M. Hewstone, & R. Brown (Eds.), *Contact and conflict in intergroup encounters* (pp. 1–44). Basil Blackwell.

Hewstone, M., Cairns, E., Voci, A., Hamberger, J., & Niens, U. (2006). Intergroup contact, forgiveness, and experience of "the troubles" in Northern Ireland. *Journal of Social Issues, 62*(1), 99–120. https://doi.org/10.1111/j.1540-4560.2006.00441.x

Imperato, C., & Mancini, T. (2022). A constructivist point of view on intergroup relations: Online intergroup contact, Dialogical Self and prejudice reduction. *Psicologia Sociale, 17*(3), 359–380.

Imperato, C., Schneider, B. H., Caricati, L., Amichai-Hamburger, Y., & Mancini, T. (2021). Allport meets internet: A meta-analytical investigation of online intergroup contact and prejudice reduction. *International Journal of Intercultural Relations, 81*, 131–141. https://doi.org/10.1016/j.ijintrel.2021.01.006

Islam, M. R., & Hewstone, M. (1993). Dimensions of contact as predictors of intergroup anxiety, perceived out-group variability, and out-group attitude: An integrative model. *Personality and Social Psychology Bulletin, 19*(6), 700–710. https://doi.org/10.1177/0146167293196005

Johnson, W. L., Rickel, J. W., & Lester, J. C. (2000). Animated pedagogical agents: Face-to-face interaction in interactive learning environments. *International Journal of Artificial Intelligence in Education, 11*(1), 47–78.

Kroh, M. (2007). Measuring left–right political orientation: The choice of response format. *Public Opinion Quarterly, 71*(2), 204–220. https://doi.org/10.1093/poq/nfm009

Kurdi, B., Banaji, M. R., & Phelps, E. A. (2019). The relationship between implicit intergroup attitudes and behavior: A meta-analysis. *Psychological Bulletin, 145*(6), 622–645. https://doi.org/10.1037/bul0000190

Maunder, R. D., White, F. A., & Verrelli, S. (2019). Modern avenues for intergroup contact: Using E-contact and intergroup emotions to reduce stereotyping and social distancing against people with schizophrenia. *Group Processes & Intergroup Relations, 22*(7), 947–963. https://doi.org/10.1177/1368430218794874

McConnell, A. R., & Leibold, J. M. (2001). Relations among the implicit association test, discriminatory behavior, and explicit measures of racial attitudes. *Journal of Experimental Social Psychology, 37*(5), 435–442. https://doi.org/10.1006/jesp.2000.1470

Montoya, A. K. (2019). Moderation analysis in two-instance repeated-measures designs: Probing methods and multiple moderator models. *Behavior Research Methods, 51*(1), 61–82. https://doi.org/10.3758/s13428-018-1088-6

Nass, C., & Moon, Y. (2000). Machines and mindlessness: Social responses to computers. *Journal of Social Issues, 56*(1), 81–103. https://doi.org/10.1111/0022-4537.00153

National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research. (1979). *The Belmont Report: Ethical principles and guidelines for the protection of human subjects of research*. U.S. Department of Health, Education, and Welfare. https://www.hhs.gov/ohrp/sites/default/files/the-belmont-report-508c_FINAL.pdf.

Nosek, B. A., Smyth, F. L., Hansen, J. J., Devos, T., Lindner, N. M., Ranganath, K. A., … Banaji, M. R. (2007). Pervasiveness and correlates of implicit attitudes and stereotypes. *European Review of Social Psychology, 18*(1), 36–88. https://doi.org/10.1080/10463280701489053

Nowak, K., & Biocca, F. (2003). The effect of the agency and virtual image of virtual humans on users' sense of presence, copresence, and social presence in virtual environments. *Presence: Teleoperators and Virtual Environments, 12*(5), 481–494. https://doi.org/10.1162/105474603322761289

Park, S. (2012). Mediated intergroup contact: Concept explication, synthesis, and application. *Mass Communication & Society, 15*(2), 136–159. https://doi.org/10.1080/15205436.2011.558804

Pereira da Costa, L., Bierwiaczonek, K., & Bianchi, M. (2024). Does digital intergroup contact reduce prejudice? A meta-analysis. *Cyberpsychology, Behavior, and Social Networking, 27*(7), 440–451. https://doi.org/10.1089/cyber.2023.0591

Pettigrew, T. F. (1998). Intergroup contact theory. *Annual Review of Psychology, 49*(1), 65–85. https://doi.org/10.1146/annurev.psych.49.1.65

Pettigrew, T. F., & Tropp, L. R. (2006). A meta-analytic test of intergroup contact theory. *Journal of Personality and Social Psychology, 90*(5), 751–783. https://doi.org/10.1037/0022-3514.90.5.751

Piñeiro-Martín, A., García-Mateo, C., Docío-Fernández, L., & López-Fernández, M. C. (2023). Ethical challenges in the development of virtual assistants powered by large language models. *Electronics, 12*(14), Article 3170. https://doi.org/10.3390/electronics12143170

Rapp, A., Curti, L., & Boldi, A. (2021). The human side of human-chatbot interaction: A systematic literature review of ten years of research on text-based chatbots. *International Journal of Human-Computer Studies, 151*, Article 102630. https://doi.org/10.1016/j.ijhcs.2021.102630

Ruane, E., Birhane, A., & Ventresque, A. (2019). Conversational AI: Social and ethical considerations. In *Proceedings of the 27th AICS conference*. CEUR Workshop Proceedings.

Sahab, S., Haqbeen, J., Hadfi, R., Ito, T., Imade, R. E., Ohnuma, S., & Hasegawa, T. (2024). E-contact facilitated by conversational agents reduces interethnic prejudice and anxiety in Afghanistan. *Communications Psychology, 2*(1), 22. https://doi.org/10.1038/s44271-024-00070-z

Schiappa, E., Gregg, P. B., & Hewes, D. E. (2005). The parasocial contact hypothesis. *Communication Monographs, 72*(1), 92–115. https://doi.org/10.1080/0363775052000342544

Schwab, K. (2024). The fourth industrial revolution: What it means, how to respond1. In *Handbook of research on strategic leadership in the fourth industrial revolution* (pp. 29–34). Edward Elgar Publishing. https://doi.org/10.4337/9781802208818.00008.

Shen, L. (2010). On a scale of state empathy during message processing. *Western Journal of Communication, 74*(5), 504–524. https://doi.org/10.1080/10570314.2010.512278

So, C., Khvan, A., & Choi, W. (2023). Natural conversations with a virtual being: How user experience with a current conversational AI model compares to expectations. *Computer Animation and Virtual Worlds, 34*(6), Article e2149. https://doi.org/10.1002/cav.2149

Song, Y., Luximon, A., & Luximon, Y. (2021). The effect of facial features on facial anthropomorphic trustworthiness in social robots. *Applied Ergonomics, 94*, Article 103420. https://doi.org/10.1016/j.apergo.2021.103420

Stephan, W. G., & Stephan, C. W. (1985). Intergroup anxiety. *Journal of Social Issues, 41*(3), 157–175. https://doi.org/10.1111/j.1540-4560.1985.tb01134.x

Tajfel, H., Billig, M. G., Bundy, R. P., & Flament, C. (1971). Social categorization and intergroup behaviour. *European Journal of Social Psychology, 1*(2), 149–178. https://doi.org/10.1002/ejsp.2420010202

Tassinari, M., Aulbach, M. B., & Jasinskaja-Lahti, I. (2022a). The use of virtual reality in studying prejudice and its reduction: A systematic review. *PLoS One, 17*(7), Article e0270748. https://doi.org/10.1371/journal.pone.0270748

Tassinari, M., Aulbach, M. B., & Jasinskaja-Lahti, I. (2022b). Investigating the influence of intergroup contact in virtual reality on empathy: An exploratory study using AltspaceVR. *Frontiers in Psychology, 12*, Article 815497. https://doi.org/10.3389/fpsyg.2021.815497

Tukachinsky, R. (2010). Para-romantic love and para-friendships: Development and assessment of a multiple-parasocial relationships scale. *American Journal of Media Psychology, 3*(1/2), 73–94.

Turner, J. C. (1981). The experimental social psychology of intergroup behaviour. In J. C. Turner (Ed.), *Intergroup behaviour* (pp. 66–102). Basil Blackwell.

Vezzali, L., Hewstone, M., Capozza, D., Giovannini, D., & Wöelfer, R. (2014). Improving intergroup relations with extended and vicarious forms of indirect contact. *European Review of Social Psychology, 25*, 314–389. https://doi.org/10.1080/10463283.2014.982948

Vezzali, L., Stathi, S., Giovannini, D., Capozza, D., & Trifiletti, E. (2015). The greatest magic of Harry Potter: Reducing prejudice. *Journal of Applied Social Psychology, 45*(2), 105–121. https://doi.org/10.1111/jasp.12279

Vivian, J., Hewstone, M., & Brown, R. (1997). Intergroup contact: Theoretical and empirical developments. In W. Stroebe, & M. Hewstone (Eds.), *Enhancing education in heterogeneous schools: Theory and application* (pp. 13–46). Sage.

Von der Pütten, A. M., Krämer, N. C., & Gratch, J. (2010). Who's there? Can a virtual agent really elicit social presence? *International Journal of Human-Computer Studies, 68*(6), 456–469. https://doi.org/10.1016/j.ijhcs.2010.02.004

White, F. A., & Abu-Rayya, H. M. (2012). A dual identity-electronic contact (DIEC) experiment promoting short- and long-term intergroup harmony. *Journal of Experimental Social Psychology, 48*(3), 597–608. https://doi.org/10.1016/j.jesp.2012.01.007

White, F. A., Harvey, L. J., Abu-Rayya, H. M., & Hisham, M. (2015). Improving intergroup relations in the internet age: A critical review. *Review of General Psychology, 19*(2), 129–139. https://doi.org/10.1037/gpr0000036

White, F. A., Maunder, R., & Verrelli, S. (2020). Text-based E-contact: Harnessing cooperative internet interactions to bridge the social and psychological divide. *European Review of Social Psychology, 31*(1), 76–119. https://doi.org/10.1080/10463283.2020.1753459

Wilder, D. A. (1984). Predictions of belief homogeneity and similarity following social categorization. *British Journal of Social Psychology, 23*(4), 323–333. https://doi.org/10.1111/j.2044-8309.1984.tb00648.x

Wong, N. C. H., Massey, Z. B., & Barbarti, J. L. (2022). Theorizing prejudice reduction via mediated intergroup contact: Extending intergroup contact theory to media contexts. *Journal of Media Psychology, 34*(2), 85–98. https://doi.org/10.1027/1864-1105/a000338

Wright, S. C., Aron, A., McLaughlin-Volpe, T., & Ropp, S. A. (1997). The extended contact effect: Knowledge of cross-group friendships and prejudice. *Journal of Personality and Social Psychology, 73*(1), 73–90. https://doi.org/10.1037/0022-3514.73.1.73

Xu, K., Nosek, B., & Greenwald, A. G. (2014). Psychology data from the race implicit association test on the project implicit demo website. *Journal of Open Psychology Data, 2*(1), Article e3. https://doi.org/10.5334/jopd.ac