Original Research Article

Recognizing and Relating to the Race/Ethnicity and Gender of Animated Pedagogical Agents

Journal of Educational Computing Research
2023, Vol. 0(0) 1–27

DOI: 10.1177/07356331231213932
journals.sagepub.com/home/jec
Sage

Fangzheng Zhao[1], Richard E. Mayer[1], Nicoletta Adamo-Villani[2], Christos Mousas[2], Minsoo Choi[2], Luchcha Lam[2], Magzhan Mukanova[2], and Klay Hauser[2]

Abstract

This study examined how well people can recognize and relate to animated pedagogical agents of varying ethnicities/races and genders. For both Study 1 (realistic-style agents) and Study 2 (cartoon-style agents), participants viewed brief video clips of virtual agents of varying racial/ethnic categories and gender types and then identified their race/ethnicity and gender and rated how human-like and likable the agent appeared. Participants were highly accurate in identifying Black and White agents but were less accurate for Asian, Indian, and Hispanic agents. Participants were accurate in recognizing gender differences. Participants rated all types of agents as moderately human-like, except for White agents. Likability ratings were lowest for White and male agents. The same pattern of results was obtained across two independent studies with different participants and different onscreen agents, which indicates that the results are not solely due to one specific set of agents. Consistent with the Media Equation Hypothesis and the Alliance Hypothesis, this work shows that people are sensitive to the race/ethnicity and gender of onscreen agents and relate to them differently. These findings

[1]University of California, Santa Barbara, CA, USA
[2]Purdue University, West Lafayette, IN, USA

Corresponding Author:
Fangzheng Zhao, Department of Psychological and Brain Sciences, University of California Santa Barbara, Santa Barbara, CA 93106, USA.
Email: fangzheng.zhao@psych.ucsb.edu

Data Availability Statement included at the end of the article

have implications for how to design animated pedagogical agents for improved multimedia learning environments in the future and serve as a crucial first step in highlighting the possibility and feasibility of incorporating diverse onscreen virtual agents into educational computer software.

Keywords
animated onscreen agents, race, ethnicity, gender, media equation theory

Introduction

Objective and Rationale

Onscreen animated pedagogical agents can be used as virtual instructors in video lectures. Given the salience of racial/ethnic and gender cues in people's perception of human instructors, we are interested in people's perceptions of onscreen virtual agents that are designed to vary in racial/ethnic characters[1] (e.g., Asian, Black, Hispanic, Indian, and White) and gender characteristics[2] (e.g., male and female). In the present study, we examine people's perceptions of 10 categories of onscreen agents: Asian male, Black male, Hispanic male, Indian male, White male, Asian female, Black female, Hispanic female, Indian female, and White female. The first goal of the present study is to examine whether people can perceive the racial/ethnic category and gender category of onscreen animated agents. The second goal is to determine whether people prefer certain categories in terms of how much they like the virtual agents and how human-like they appear. The final goal is to determine whether the same pattern of results is obtained across two studies involving different sets of onscreen agents and different participants.

The rationale behind conducting this study lies in the potential impact of racial/ethnic and gender categories on learners' affective, social, and cognitive processes during video lectures. By employing animated virtual agents, we can create identical video lectures that vary solely in terms of the racial/ethnic category and gender category of the onscreen virtual instructor. This approach substantially enhances the diversity and representation of the teaching cohort in online education. Unlike human instructors who are constrained by their inherent racial/ethnic and gender identities, virtual agents afford us the flexibility to easily modify their appearances and adapt their voice tones, allowing for a more inclusive and diverse teaching environment for students from different backgrounds. Therefore, as a first step, we seek to evaluate how perceive the virtual agents generated by recent technology. Specifically, we aim to investigate whether people are able recognize the racial/ethnic and gender category of onscreen agents and their affective response (e.g., ratings of likeability) and social response (e.g., ratings of human-likeness) to them.

In short, our objective is to determine whether participants can perceive the race/ethnicity and the gender of the animated onscreen agents accurately, and the extent to which they like the agents and feel the agents were human-like. Beyond these immediate objectives, our research carries substantial theoretical significance by shedding light on whether racial and gender stereotypes apply to virtual agents, suggesting they treat onscreen agents like real humans as proposed by Media Equation Theory. Additionally, this study has practical implications for the enhancement of animated pedagogical agent design and underscores the potential of utilizing animated pedagogical agents within multimedia platforms to provide a more inclusive and diverse learning environment.

Theoretical Background

Media Equation Theory

The Media Equation Theory, proposed by Reeves and Nass (1996), posits that humans can naturally treat computers and other media as if they were real, social entities. This phenomenon of anthropomorphism arises from the inherent social and cognitive tendencies ingrained in human beings' evolved brains. Consequently, under appropriate conditions, individuals can automatically accept artificial onscreen agents and environments as if they are real, leading them to engage with those technologies in ways similar to how they interact with other real humans. Numerous empirical studies (Lawson et al., 2021a, 2021b, 2021c; Lawson & Mayer, 2021; Nass et al., 1997; Nass & Steuer, 1993; Zhao & Mayer, 2023) have provided evidence supporting the Media Equation Theory. These studies have picked various aspects of how people respond to each other or to the natural environments, such as politeness, personalities, gender, and emotional responses, and investigated whether such the patterns of human responses in natural environment could replicate in human-computer or human-technology interactions.

For instance, one study examining polite interactions with computers employed a comparison between two participant groups evaluating a computer's performance in teaching them a lesson (Nass & Steuer, 1993). Participants evaluated the computer's performance either on the same computer or on a different computer in another room. Surprisingly, participants who interacted with the same computer provided significantly more positive responses than those interacting with a different computer. Subsequent research using both text-based and voice-based interactions with computers replicated these findings, revealing that individuals exhibit politeness towards computers regardless of the interaction medium. In short, the participants treated computers like real people by not wanting to hurt their feelings.

Another study explored gender stereotypes by assigning stereotypical topics to female and male voices in computer-based interactions (Nass et al., 1997). The results indicated that people are capable of detecting gender differences and applying stereotypes typically associated with real human beings based on subtle cues displayed on

computers, such as voices. Such the evidence strongly suggests that whether through plain text or voice alone, individuals have been shown to engage in social interactions with technologies such as a computer and such the social interactions can easily be triggered.

In addition, an fMRI study, aligning with the principles of the Media Equation Theory, has contributed valuable insights into how people extend their emotional responses from human interactions to interactions with robots (Rosenthal-Von Der Putten et al., 2014). In a study involving 14 participants, video stimuli were employed to depict interactions with human subjects, a baby camarasaurus-shaped entertainment robot, and inanimate subjects, each exhibiting either affectionate or violent behavior. The findings, based on a combination of self-reported emotional states and fMRI data, illustrated that participants exhibited emotional responses when exposed to scenarios featuring expressions of affection or violence involving both humans and robots. Interestingly, while distinct neural activation patterns were discerned in the context of abusive behavior scenarios which indicated that participants showed more emotional responses for real human, the neural activity was similar in response to affectionate interactions with both humans and robots, despite the robot's non-humanoid form. Furthermore, another study, which analyzed pupillary responses to both robotic and human emotions, revealed a noticeable similarity in the patterns of pupil responses across various emotions between the robot and human stimuli, which confirmed the Media Equation hypothesis (Reuten et al., 2018).

While the original concepts and research associated with the Media Equation Theory date back several decades, there have been many more recent studies that continue to contribute to and expand upon this theory using more advanced technology related to virtual agents and even robots (Domagk, 2008, 2010; Sträfling et al., 2010; Nowak & Rauh, 2005; Reuten et al., 2018). For example, a recent study examined human interactions with the intelligent virtual agent, Siri, shedding light on the development of social connections and relational dynamics with Siri. Their findings demonstrated that increased trust in cognitive aspects aligned with successful task completion facilitated by Siri, aligning with the principles of the Media Equation Theory.

Furthermore, according to a review chapter written by Kramer et al. (2015), several other studies have explored the specific features of these agents that influence human interactions (Domagk, 2008), which emphasized that a likable appearance contributes to more positive outcomes (Domagk, 2010). To explore this further, a study by Sträfling et al. (2010) compared the influence of a cartoon-like rabbit character to a realistic anthropomorphic agent as a virtual tutor. The results revealed a preference for the rabbit-like agent than the realistic anthropomorphic agent, while another study focusing on credibility found that more anthropomorphic characters tend to be perceived as more credible (Nowak & Rauh, 2005). In addition, Reuten et al. (2018) conducted a study assessing pupillary responses to robots and found that increased human-likeness was associated with increased difficulty in recognizing emotions.

In summary, in light of the numerous previous research studies about the Media Equation Theory and the advancements in technology related to virtual agents, there

has been a lot of evidence supporting Media Equation Theory, affirming that individuals are capable of forming social connections with technological entities, including computers, robots, and virtual agents. However, the precise features of the agents, such as their degree of anthropomorphism, gender categorizations, and other dimensions, warrant further investigation to determine their efficacy and specific impacts on interactions between humans and technology. Building upon these findings and research gaps, the present study aims to contribute to the Media Equation Theory by examining how people perceive and respond to on-screen virtual agents with varying racial/ethnic and gender categories, in the absence of any accompanying voices. The virtual agents utilized in this research were created using the most recent 3D modeling technologies, with two distinct styles: realistic and cartoonish. By focusing solely on visual cues, particularly racial/ethnic and gender types, this investigation seeks to enhance our understanding of whether individuals can accurately perceive different agent categories and react to them differently. Furthermore, as for more global interest, this investigation seeks to offer valuable insights on the potential of using diverse virtual agents in multimedia educational contexts, ultimately fostering a more inclusive and diverse learning environment beneficial to students from different backgrounds.

Alliance Hypothesis

The Alliance Hypothesis posits that humans have a tendency to form alliances with individuals who share similar physical characteristics, ultimately facilitating cooperation and collective success (Taylor et al., 1978). This hypothesis indicates that racial categorization is a byproduct of the psychology of alliance (Kurzban et al., 2001; Pietraszewski, 2021; Taylor et al., 1978), which highlights the role of race as a visual cue that influences alliance formation, subsequently impacting social behavior and group dynamics. By delving into the perception and categorization of race within human society, this hypothesis seeks to shed light on how the concept of race came into existence and how race influences cooperative behaviors, social preferences, and intergroup relations. Rooted in evolutionary theories such as kin selection and reciprocal altruism (Michod, 1982; Eberhard, 1975), the Alliance Hypothesis argues that humans possess an inherent inclination to form alliances with genetically similar individuals, as it enhances the likelihood of mutual support and cooperation.

Numerous studies have extensively examined and provided support for the Alliance Hypothesis, revealing that individuals reliably form groups characterized by sharp boundaries. These groups are often marked based on arbitrary, superficial, and context-dependent indicators, such as skin tones, gender types, clothing styles, speech patterns, and behaviors. For example, in a study conducted by Taylor et al. (1978), participants, who were equally divided by sex, were tasked with observing a group discussion while being shown pictures of the participants as they spoke. Half of the participants were given control instructions to observe the group as a whole, while the other half were given memory instructions to match each suggestion with a picture of the speaker to investigate when categorization by race (Experiment 1) or sex (Experiment 2) occurs

and whether memory instructions would affect the recall of race-related or sex-related information. After the discussion, participants were asked to match suggestions to the participants (i.e., the "Who Said What?" task). The results showed that individuals tend to categorize others based on race and sex, leading to the more within-category attribution errors than more between-category attributions errors. That is, people tend to minimize the within-group differences and amplify the between-group differences.

Numerous prior studies have demonstrated that the inclination to categorize individuals into distinct identity groups involves computational processes that appear to operate automatically and involuntarily (Brewer, 1979; Tajfel et al., 1971; Kurzban et al., 2001). When encountering a new person, the human mind is triggered to activate three fundamental dimensions—namely, race, gender, and age (Messick & Mackie, 1989; Hamilton et al., 1994; Fiske & Neuberg, 1990). These dimensions are encoded automatically and mandatorily, irrespective of the social context (Hewstone et al., 1991; Stangor et al., 1992). This encoding can occur independently of any additional individual-specific information. For instance, individuals may recall that a new acquaintance is a young white woman without retaining any further details about her. In one classic study conducted by Kurzban et al. (2001), the findings indicated that when there were no additional visual cues specifying the group membership, participants tended to automatically categorize individuals into groups based on races and gender types. However, when additional visual cues about coalition were provided, the inclination of participants to categorize individuals by race significantly decreased, while gender types consistently remained more strongly encoded than coalition membership.

In addition, some more recent studies also revealed how people were able to automatically identify and differentiate various racial/ethnic categories and gender types (Cosmides et al., 2003; Pietraszewski, 2009, 2016). For example, in Pietraszewski's study (2021), participants encountered various scenarios involving basketball teams. In the baseline situation, they could see players' races but not their team affiliations. In the moderate scenario, participants inferred team memberships based on shirt colors, leading to confusion. In the strong scenario, both races and team memberships were evident. As the results, in the moderate condition, participants showed more confusion in categorizing team memberships. In the baseline situation, where no hints about membership were given, participants automatically associated race with alliances. In the strong situation, participants prioritized team categorization when memberships were clear, temporarily suppressing the automatic race categorization. However, this suppression effect was less consistent when players differed in gender types rather than races, highlighting gender's robust influence within the Alliance Hypothesis mechanism.

In conclusion, the existing evidence encourages the idea that people can effectively differentiate between various racial and gender types, automatically perceiving them as distinct alliances.

However, previous studies have primarily examined how individuals categorize racial and gender types among human groups, leaving a gap in research concerning the application of the Alliance Hypothesis to non-human entities, specifically virtual

agents. Therefore, in this present study, we aimed to also explore whether the inherent automatic ability to identify and categorize groups based on racial/racial or gender types could extend to virtual agents displaying distinct visual features indicative of various racial and gender categories. As the global interest of this study, such the investigation presented an opportunity not only to validate the advancements in current 3D virtual agent technologies but also to underscore the potential of employing agents featuring diverse racial/ethnic and gender characteristics to align with students' backgrounds and preferences. This approach holds the promise of fostering a more inclusive and representative learning environment.

Racial Stereotypes to Artificial Intelligent Agents

In addition to the previous studies investigating the Alliance Hypothesis mainly on observing real human pictures and voices, there was one recent study that investigated the ascription of racial stereotypes to artificially intelligent (AI) agents (Davis et al., 2023), building upon the Stereotype Content Model (Fiske et al., 2002) and the Humanness-Value-Loyalty Framework (Belanche et al., 2021). The Stereotype Content Model posits that stereotypes are characterized by two fundamental dimensions: warmth (reflecting perceived friendliness, trustworthiness) and competence (reflecting perceived intelligence, capability). The Humanness-Value-Loyalty Framework suggests that visual cues related to AI robots' behavior inform customers' perceptions of the robot's humanness, competence, and warmth in service contexts. In a recent negotiation study (Davis et al., 2023), participants interacted with AI bots representing different racial groups (Black, White, or Asian) in a simulation of an online home-sharing platform. They aimed to negotiate the lowest price for temporary housing. Surprisingly, participants rated Black AI bots more positively than White or Asian bots, perceiving them as more competent and human-like. These findings highlight the role of race in shaping consumer responses to AI agents, suggesting distinctions between human-bot and human-human interactions.

In summary, this research offers preliminary evidence that individuals can accurately discern and respond differently to animated onscreen agents of varying racial categories. However, it's important to acknowledge that the interactions between humans and cartoonish onscreen agents displayed some differences from human-to-human interactions, challenging prior findings that support the Media Equation Theory (Davis et al., 2023). Moreover, the AI bots in Davis et al.'s study (2023) did not possess a realistic style or incorporate diverse gender types, which might introduce some limitations. Therefore, further investigation in the present study aimed to fill this research gap by further investigating how people perceive and react to animated onscreen agents representing diverse racial/ethnic and gender categories, encompassing both realistic and cartoonish styles, and whether their responses align with those observed in interactions with real humans.

Theory and Predictions

In accordance with the Alliance Hypothesis, racial/ethnic and gender categories are inherently salient human features, prompting individuals to automatically identify and categorize these attributes. Therefore, it is worth exploring whether people can similarly discern these features in animated onscreen agents. In addition, according to the Media Equation Theory (Reeves & Nass, 1997), people easily can be induced to treat technology-based beings the same way they treat real humans. Accordingly, people should be able to recognize the racial/ethnic category and gender category of onscreen animated agents. Also, people should be able to feel affective responses to different categories of onscreen animated agents, such as how much they like them. Furthermore, according to the Stereotype Content Model, people focus on the dimensions of warmth and competence in making sense of individuals or groups (Fiske, 2018). Our ratings of likability are most closely related to the warmth dimension.

To investigate these issues, we conducted two independent studies in which people viewed short video clips of various onscreen agents and for each clip indicated the agent's racial/ethnic category, gender, level of human-likeness, and level of likability. In the current study, we aim to address the following research questions:

***Research Question 1*:** Are people able to recognize the racial/ethnic category of onscreen agents? In particular, we are interested in the degree to which people can recognize the racial/ethnic category for agents intended to be Asian, Black, Hispanic, Indian, or White.

***Research Question 2*:** Are people able to recognize the gender category of onscreen agents? In particular, we are interested in the degree to which people can recognize the gender category for agents intended to be male or female.

***Research Question 3*:** What level of human-likeness do people report for each of the 10 categories of onscreen agents? In particular, we are interested in the ratings of human-likeness for each racial/ethnic and gender category in the studies.

***Research Question 4*:** How much do people like each of the 10 categories of onscreen agents? In particular, we are interested in the ratings of likability for each of the racial/ethnic and gender categories in the studies.

***Research Question 5*:** Is the pattern of results similar across two studies using different onscreen agents in each category? In particular, we want to know if people can recognize each racial/ethnic category and each gender category equivalently across different sets of onscreen agents. We also want to know whether the pattern of ratings on human-likeness and likability for each racial/ethnic and gender category are equivalent across studies with different sets of onscreen agents. Replication across studies is helpful in establishing that the results do not depend on a specific set of agents.

Method

Study 1 (Realistic Agents)

Study 1 employed a within-subject design study, in which participants were asked to watch short video clips of each of 37 animated pedagogical agents twice. After watching each video clip, participants took a short questionnaire in which they were asked to identify the racial/ethnic category of the agent, and rate the agent on gender characteristics, how human-like the agent appeared, and how much they like the agent.

Participants. The participants in Study 1 were 82 undergraduate students from the University of California, Santa Barbara (UCSB). The participants were recruited from the Psychology Subject Pool and their participation in the study fulfilled a course requirement. The mean age was 19.05 years (SD = 1.24). Out of the 82 participants, 46 identified as female, 33 identified as male, 2 identified as non-binary, and 1 chose not to disclose their gender. In terms of racial/ethnic category, 24 participants identified as Asian, 4 as Black/African American, 24 as Hispanic/Latino, 6 as Indian, 22 as White/Caucasian, and 3 as other. UCSB is a Hispanic-Serving Institution (HSI) and an Asian American and Native American Pacific Islander-Serving Institution (AANAPISI), which means that the undergraduate enrollment is at least 25% Hispanic and at least 10% Asian American and Native American Pacific Islander.

Design. Using a within-subjects design, all 82 participants watched 37 different types of agent videos twice, with the order of presentation randomized within each block. This resulted in a total of 74 videos being watched by each participant. The dependent measures for all participants included accuracy in identifying the race/ethnicity and gender of the agents, ratings of the agents' level of human-like behavior, and ratings of the agents' likability.

Materials. We constructed an online Qualtrics survey that consisted of a consent form, a set of 37 short video chips depicting animated pedagogical agents with each clip presented twice, a 4-item survey after each clip, a post-questionnaire, and a debriefing statement.

Virtual Agents. A collection of 37 pedagogical agents were designed in a realistic style. This process encompassed the manipulation of three fundamental attributes: gender, skin color, and hair color, utilizing Character Creator 4, a software resource provided by Reallusion. Within its extensive repository, encompassing both its default library and the supplementary Working Class Heroes collection available in its contents store, a wide array of virtual character assets was accessible, encompassing aspects such as skin color, body types, and clothing attire. Notably, these assets included six distinct skin tones: Asian pale, black, brown, Caucasian white, tan, and yellow. Moreover, Character Creator facilitated further customization through its suite of

editing tools, allowing for alterations to clothing attire and hair color, thereby enabling the introduction of considerable variation among the virtual characters. It is noteworthy that certain elements, such as clothing and body shapes, were specifically curated by the character designer in our research team to align with the research objectives and aesthetics. Importantly, we carefully selected and designed specific features of the virtual agents to ensure that students could effectively differentiate between the various races/ethnicities and genders of the agents, underscoring our commitment to fostering meaningful distinctions beyond just presumed profiles.

Video Clips. Each type of agent was represented by a 3-s video clip. The 37 different types of videos included 10 categories of agents: Asian male ($n = 4$), Black male ($n = 3$), Hispanic male ($n = 4$), Indian male (n dd = 4), White male ($n = 3$), Asian female ($n = 3$), Black female ($n = 5$), Hispanic female ($n = 3$), Indian female ($n = 4$), While female ($n = 4$). Screenshot examples of the agents are shown in Figure 1. All of the video clips can be viewed through this link: (https://osf.io/zy2tm/?view_only=9e6e67ec00434d15aa052be8d66ccbc8). Since each video clip and accompanying questionnaire was presented twice in random order, participants watched 74 video clips in total.

Post-video Questionnaire. Since our goal is to understand how people perceive virtual agents of different racial/ethnic categories and genders, this entails accessing both the accuracy of identifying various types of the agents and gauging the perceived level of human-likeness and favorability towards those agents. To achieve this, we administered four questions following each video clip. The first set of questions aimed to establish whether participants could identify the race/ethnicity and gender of these agents accurately. Subsequently, we explored whether participants perceived some specific types of agents as more human-like and likable. Therefore, the first question asked the participant to identify the racial/ethnic category of the agent presented in the video clip (i.e., "Which ethnicity do you think this agent is?"), with five answer options: Asian, Black/African American, Hispanic/Latino, Indian, and White/Caucasian. The goal of this question is to measure how accurate participants could identify the race/ethnicity types of the various virtual agents. Then, participants were asked to rate how masculine/feminine the agent appeared on a seven-point scale ranging from "Very feminine" (1) to "Very masculine" (7), in response to the item "How masculine/feminine do you think this agent is?". In this approach, we were able to record how accurately participants could identify the gender types of the various virtual agents. After that, to investigate the degree of perceived human-likeness attributed to the virtual agents by participants, participants were asked to provide ratings on a 7-point scale ranging from "Extremely non-human" (1) to "Extremely human-like" (7), in response to the item "How human-like do you think this agent is?". Finally, participants rated how much they liked the agent on a 7-point scale ranging from "Dislike a great deal" (1) to "Like a great deal" (7), in response to the item "How much do you like this agent?". The goal of this question is to measure the level of perceived likeability of the various virtual agents by participants, which may be an important feature for instructional videos.

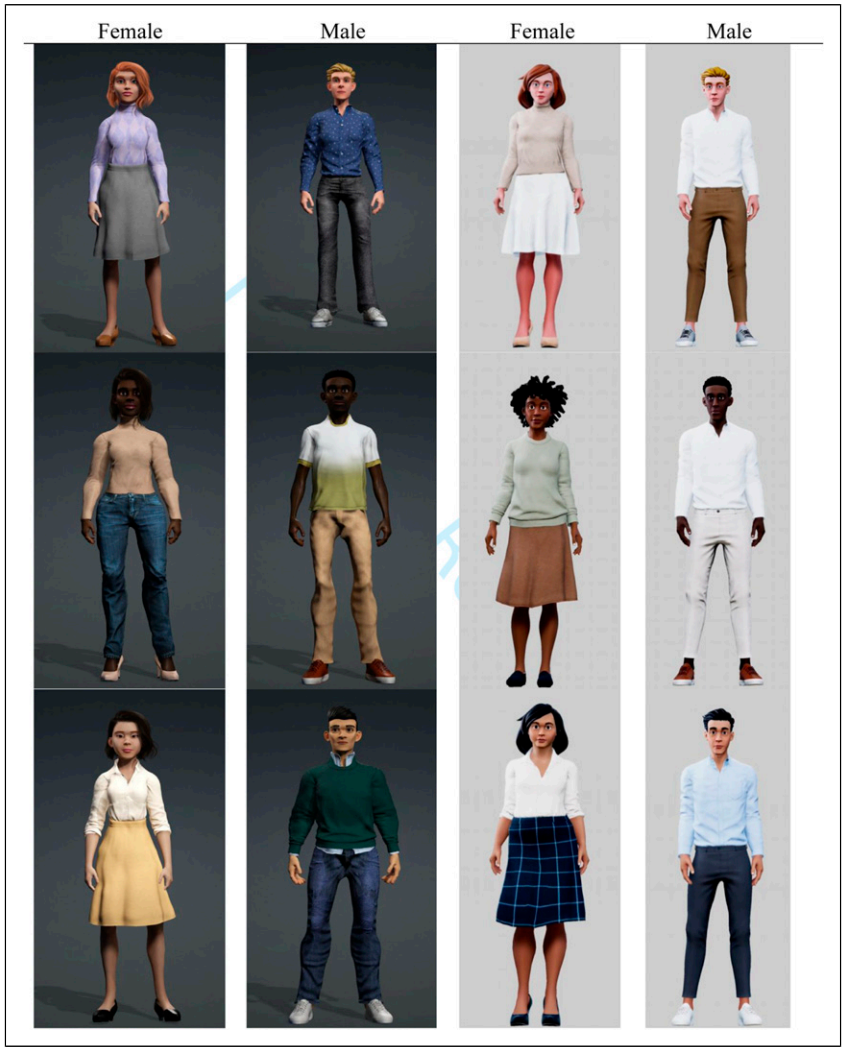

Figure 1. Example Images of Realistic versus, Example Images of Realistic versus Cartoonish Agents by Gender & Racial/ethnic Types.
Note. The images on the first row represent examples of White agents, the images on second row represent examples of Black agents, and the images on the third row represent examples of Asian Agents.

Demographics. After all video clips and post-video questionnaires, participants were asked to provide demographic information including their age, gender (male, female, or non-binary), and racial/ethnic category (Asian, Black/African American, Hispanic/Latino, Indian, White/Caucasian, or Other).

Consent and Debriefing. Before the study there was a consent form that described the participants' rights and after the study there was a debriefing statement that described the study.

Procedure. The study was conducted online through the UCSB SONA participant recruitment system. Before beginning the study, participants were provided with a Qualtrics link and asked to read and agree to the consent form. Once the consent form was signed, participants watched a total of 74 3-s videos, each featuring a different type of 3D animated pedagogical agent. The first block of 37 short video clips was presented to participants in random order; then, the same set of 37 clips was presented again in random order. After each video, a brief survey was provided to participants. Once all the videos had been watched and all survey questions answered, participants were asked to complete a brief demographic questionnaire, provided with a debriefing form, and thanked for their participation. We followed guidelines for research with human subjects and obtained IRB approval (Protocol Number 176-23-0016).

Study 2 (Cartoonish Agents)

Study 2 served as a replication of study 1, except that the more realistic agents in Study 1 were replaced with more cartoon-style agents in study 2 and the participants were different.

Participants. The participants were 86 undergraduate students at the University of California, Santa Barbara from the same subject pool as in Study 1. Their mean age of was 19.08 years (SD = 1.03). Of the participants, 53 identified as female, 32 as male, and 1 as non-binary. In terms of ethnicity, 18 participants identified as Asian, 6 as Black/African American, 26 as Hispanic/Latino, 35 as White/Caucasian, and 1 as Indian.

Design. The design of Study 2 was nearly identical to that of Study 1, with the only difference being the replacement of the 3-s video clips featuring realistic agents with 3-s video clips featuring cartoonish agents. Specifically, 86 participants in study 2 watched all 36 types of agent videos twice, presented in a random order, resulting in a total of 72 videos watched per participant. The dependent measures used in Study 2 were the same as those used in Study 1.

Materials. As in Study 1, each pedagogical agent in Study 2 was presented in a 3-s video clip. However, in contrast to Study 1, the agents in Study 2 were designed to be more cartoon-like and less realistic in style. These agents were also created using the same tool, Reallusion's Character Creator 4, as in Study 1. The new version of the agents used in Study 2 can be referred to as cartoonish agents due to their design, featuring larger eyes, enlarged heads, and slightly exaggerated body characteristics compared to the version of the agents used in Study 1. We used the Working Class

Heroes collection to generate virtual characters with cartoon styles. There was a total of 36 different types of videos in Study 2, featuring the same five different races/ethnicities and same two gender types as in Study 1. More specifically, there were three different types for Asian female agents, four for Asian male agents, five for Black female agents, three for Black male agents, three for Hispanic female agents, four for Hispanic male agents, four for Indian female agents, three for Indian male agents, three for White female agents, and four for White male agents. Screenshot examples of the agents are shown in Figure 1. All of the video clips in Study 2 can be viewed through this link: (https://osf.io/zy2tm/?view_only=9e6e67ec00434d15aa052be8d66ccbc8).

The short questionnaire after each video and the demographic survey at the end of the study were the same as in Study 1.

Procedure. The procedure was the same as Study 1.

Results

Research Question 1. How Well Did Participants Recognize the Race/Ethnicity of Agents?

Table 1 shows the mean accuracy scores (and standard deviations) by race/ethnicity in the two studies.

We conducted a 5 (agent race/ethnicity: Asian, Black, Hispanic, Indian, White) x 2 (study: 1, 2) analysis of variance (ANOVA). There was a significant difference among the five racial/ethnic categories on accuracy score, $F(4, 57) = 27.85$, $p < .001$. Further pairwise comparison analysis with the Tukey HSD indicated that the participants were best at identifying the White agents (M = .93, SD = .08) and Black agents (M = .91, SD = .11), but were significantly less accurate at identifying Asian (M = .55, SD = .34), Indian (M = .38, SD = .27), and Hispanic agents (M = .23, SD = .19) agents (all ps < .001). Although accuracy scores for Asian agents were significantly lower than Black and white agents, they were significantly higher than for Hispanic agents ($p = .003$). Accuracy scores were not significantly different between White and Black agents, between Asian and Indian agents, nor between Hispanic and Indian agents.[3] There was no significant difference in accuracy scores between study 1 and study 2, $F(1, 57) = .05$, $p = .83$, and no significant interaction between racial/ethnic types and study, $F(4, 57) = 1.67$, $p = .17$. These results indicate that the identification accuracy for the five race/ethnicity types of agents was best for White and Black agents and did not differ between Study 1 and Study 2.

Research Question 2. How Well Did Participants Recognize the Gender of Agents?

Table 2 shows the mean femininity/masculinity rating for female and male agents in the two studies. We conducted a 2 (gender type: female, male) x 2 (study: 1, 2) ANOVA.

Table 1. Means and Standard Deviations of Ethnicity Identification Accuracy by Agent Racial/Ethnic Types and Gender Types in Study 1 and 2.

	White		Indian		Hispanic		Black		Asian		Total	
	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD
Study 1 – Realistic version												
Male Agent	.99	.01	.32	.34	.24	.33	.85	.14	.75	.27	.61	.38
Female Agent	.80	.03	.36	.03	.18	.08	.94	.02	.66	.43	.59	.33
Total	**.89**	.10	.34*	.24	.21*	.22	**.90**	.10	**.71**	.32		
Study 2 – Cartoonish version												
Male Agent	.96	.01	.38	.38	.24	.12	.95	.04	.34	.35	.57	.39
Female Agent	.97	.01	.45	.30	.27	.27	.90	.16	.54	.28	.65	.34
Total	**.97**	**.01**	.42*	.30	.25*	.17	**.92**	.12	.42*	.32		
Study 1 & 2												
Male Agent	.98	.01	.35	.33	.24	.21	.90	.11	.51	.37	.59	.38
Female Agent	.88	.09	.41	.22	.23	.18	.92	.12	.60	.33	.62	.33
Total	**.93**	.08	.38*	.27	.23*	.19	**.91**	.11	.55*	.34		

Note. The row named "Total" included the average scores obtained by combining the identification accuracy of ethnicity for both the female and male agents. The column named "Total" obtained by combining the identification accuracy of ethnicity for all the five ethnicity types of agents. Asterisk(*) represents significant difference from the bolded condition.

Table 2. Means and Standard Deviations of Femininity/Masculinity Ratings by Agent Racial/Ethnic Types and Gender Types in Study 1 and 2.

	White		Indian		Hispanic		black		Asian		Total	
	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD
Study 1 – Realistic version												
Male Agent	5.56	.10	6.07	.31	5.38	.53	5.99	.27	5.14	.33	5.66	.47
Female Agent	1.99	.31	2.11	.26	2.06	.41	2.18	.18	2.00	.24	2.07*	.26
Study 2 – Cartoonish version												
Male Agent	5.54	.09	5.97	.14	5.51	.25	5.86	.12	5.74	.20	5.70	.24
Female Agent	1.93	.20	2.10	.35	2.28	.37	2.47	.59	1.91	.12	2.17*	.42
Study 1 & 2												
Male Agent	5.55	.09	6.03	.24	5.45	.36	5.93	.20	5.48	.40	5.58	.35
Female Agent	1.96	.24	2.10	.29	2.17	.37	2.36	.48	1.96	.18	2.13*	.35

Note. Higher scores mean participants rate the agent to be more masculine, while lower scores mean participants rate the agent to be more feminine. The ratings for non-binary agents were not reported in this table. The column named "Total" included the average scores obtained by combining the ratings of masculinity/femininity level for all the five ethnicity types of agents. Asterisk(*) represents significant difference from the bolded condition.

There was a significant difference in femininity/masculinity ratings for female and male agents, $F(1, 63) = 1620.73$, $p < .001$, indicating that participants rated female agents as more feminine (M = 2.13; SD = .35) and male agents as more masculine (M = 5.68, SD = .36). There was no significant difference between Study 1 and Study 2 on femininity/masculinity ratings, $F(1, 63) = .74$, $p = .39$, and no significant interaction between gender types and study, $F(1, 63) = .10$, $p = .75$. These results indicate that accuracy in identifying female and male agents was high and did not differ between Study 1 and Study 2.

Research Question 3. How Did Participants Rate the Human-Likeness of the Different Types of Agents?

Table 3 shows the mean human-likeness rating (and standard deviation) for each type of agent. We conducted a 5 (racial/ethnicity type: Asian, Black, Hispanic, Indian, White) x 2 (gender: female, male) x 2 (study: 1, 2) ANOVA on human-likeness ratings. There was a significant difference in human-likeness ratings among the five racial/ethnic types of agents, $F(4, 47) = 5.45$, $p = .001$. Further pairwise comparison analysis with the Tukey HSD indicated that the White agents (M = 4.75, SD = .33) were rated to be least human-like and received human-likeness ratings significantly lower than the Black agents (M = 5.05, SD = .18, $p = .02$), Asian agents (M = 5.04, SD = .21, $p = .02$), and Indian agents (M = 5.17, SD = .18, $p < .001$). The difference on human-likeness ratings

Table 3. Means and Standard Deviations of Human-like Ratings by Agent Racial/Ethnic Types and Gender Types in Study 1 and 2.

	White		Indian		Hispanic		black		Asian		Total	
	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD
Study 1 – Realistic version												
Male Agent	4.74	.61	5.22	.27	4.97	.17	5.07	.23	4.83	.07	4.98	.33
Female Agent	4.83	.19	5.17	.10	5.01	.16	5.09	.09	5.10	.25	5.04	.19
Total	4.78	.41	5.20	.20	4.99	.15	5.08	.17	4.96	.22	-	-
Study 2 – Cartoonish version												
Male Agent	4.63	.27	5.24	.15	4.99	.23	4.95	.29	5.09	.21	4.97	.29
Female Agent	4.83	.29	5.06	.13	5.01	.16	5.07	.15	5.13	.24	5.03	.20
Total	4.72	.28	5.14	.16	5.00	.19	5.02	.20	5.11	.20	-	-
Study 1 & 2												
Male Agent	4.68	.40	5.23	.21	4.99	.19	5.01	.25	4.98	.21	4.98	.31
Female Agent	4.83	.22	5.11	.12	5.01	.14	5.07	.12	5.11	.22	5.03	.19
Total	4.75	.33	5.17*	.18	5.00	.16	5.05*	.18	5.04*	.21	-	-

Note. The row named "Total" included the average scores obtained by combining the rating of human-like level for both the female and male agents. The column named "Total" included the average scores obtained by combining the ratings of human-like level for all the five ethnicity types of agents.

between Hispanic agents and White agents did not reach statistical significance (M = 5.00, SD = .16, p = .07).[4]

There were no other significant differences. Specifically, there was no significant difference between female and male agents, F(1, 47) = .91, p = .35 on human-likeness ratings; no significant difference between Study 1 and Study 2 on human-likeness ratings, F(1, 47) = .002, p = .97; no significant interaction between racial/ethnic category and gender category, F(4, 47) = .70, p = .60; no significant interaction between ethnicity/racial category and study, F(4, 47) = .46, p = .76; no significant interaction between gender category and study, F(1, 47) = .12, p = .74; and no interaction among ethnicity/racial category, gender category, and study, F(4, 47) = .32, p = .87.

Overall, we conclude that all types of agents, regardless of the styles of the agents (i.e., realistic or cartoonish), were equivalent in being rated as moderately human-like across both studies, although White agents were rated significantly lower in human-likeness than other racial/ethnic categories.

Research Question 4. How Did Participants Rate the Likability of Different Types of Agents?

Table 4 shows the mean likability rating (and standard deviation) for each type of agent in both studies. We conducted a 5 (racial/ethnic type: Asian, Black, Hispanic, Indian, White) x 2 (gender: female, male) x 2 (study: 1, 2) ANOVA on likability ratings. There

Table 4. Means and Standard Deviations of Likability Ratings by Agent Racial/Ethnic Types and Gender Types in Study 1 and 2.

	White		Indian		Hispanic		black		Asian		Total	
	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD
Study 1 – Realistic version												
Male Agent	4.24	.34	4,67	.26	4.38	.17	4,71	.26	4.39	.06	4.99	.28
Female Agent	4.39	.29	4.78	.20	4.71	.18	4.81	.12	4.80	.18	4.70*	.24
Total	4.31	.30	4.72	.23	4.54	.24	4.76	.19	4.60	.25	-	-
Study 2 – Cartoonish version												
Male Agent	3.80	.23	4.37	.10	4.20	.28	4.30	.19	4.36	.17	4.19	.29
Female Agent	4.05	.28	4.48	.14	4.35	.15	4.55	.15	4.58	.08	4.42*	.24
Total	3.90	.26	4.43*	.12	4.26	.23	4.46*	.20	4.45*	.17	-	-
Study 1 & 2												
Male Agent	3.98	.35	4.55	.25	4.27	.24	4.50	.30	4.37	.13	4.33	.32
Female Agent	4.22	.32	4.61	.22	4.53	.25	4.65	.19	4.69	.17	4.55*	.27
Total	4.09	.34	4.58*	.23	4.39*	.27	4.59*	.24	4.52*	.22	-	-

Note. The row named "Total" included the average scores obtained by combining the rating of likability level for both the female and male agents. The column named "Total" included the average scores obtained by combining the ratings of likability level for all the five ethnicity types of agents. Asterisk(*) represents a significant difference from the bolded condition.

was a significant difference among the five ethnicity/racial types of agents on likability ratings, $F(4, 47) = 12.02$, $p < .001$. The pairwise comparison analysis with the Tukey HSD indicated that White agents ($M = 4.09$, $SD = .34$) were least liked compared to Black agents ($M = 4.59$, $SD = .24$, $p < .001$), Asian agents ($M = 4.52$, $SD = .22$, $p < .001$), Indian agents ($M = 4.58$, $SD = .23$, $p < .001$), and Hispanic agents ($M = 4.39$, $SD = .27$, $p = .005$). The likability ratings for agents' various racial/ethnicity types, except White agents, were not significantly different from each other.[5]

Additionally, there was a significant difference in likability ratings between female and male agents, $F(1, 47) = 16.76$, $p < .001$, indicating that female agents ($M = 4.55$, $SD = .27$) were more likable than male agents ($M = 4.33$, $SD = .32$).

Finally, there was a significant main effect of study, $F(1, 47) = 31.60$, $p < .001$, indicating that the realistic agents in Study 1 were better liked ($M = 4.59$, $SD = .28$) than the cartoonish agents in Study 2 ($M = 4.31$, $SD = .29$). No significant interactions were found, indicating that the pattern of differences between the five racial/ethnic types and the two gender types investigated in this work were consistent across the two studies. We conclude that agents that were least liked were White and male.

Research Question 5. Is the Pattern of Results Similar Across Two Studies Using Different Onscreen Agents in Each Category?

The pattern of differences among the Asian, Black, Hispanic, Indian, and White agents on each of the four measures was equivalent across the two studies, as is indicated by the lack of significant interaction involving study reported in the foregoing sections. The pattern of differences between female and male agents on each of the four measures was equivalent across the two studies, as is indicated by the lack of significant interaction involving study reported in the foregoing sections. We conclude that the pattern of results is not attributable to one specific set of agents.

Supplementary Analysis

As noted in the method section, each agent appeared and was rated twice. To assess the within-subject consistency of participant responses we conducted Pearson correlations on the score for each onscreen agent for each of the four rating items on the first and second time they appeared. In Study 1, the median test-retest reliability was high and significant for race/ethnicity accuracy score, $r = 1.00$, $p < .001$; femininity/masculinity rating, $r = .99$, $p < .001$; human-like rating, $r = .74$, $p < .01$; and likability rating, $r = .71$, $p < .01$. In Study 2, the median test-retest reliability was high and significant for race/ethnicity accuracy score, $r = .99$, $p < .001$; femininity/masculinity rating, $r = .79$, $p < .001$; human-like rating, $r = .71$ $p < .01$; and likability rating, $r = .66$, $p < .01$. We conclude that participants were consistent in their responses to each agent.

In Study 1 and Study 2, different participants rated different onscreen agents that represented the same 10 categories: Asian male, Asian female, Black male, Black female, Hispanic male, Hispanic female, Indian male, Indian female, White male,

White female. To assess the between-subjects consistency of responding to each of the four rating items we conducted Pearson correlations between Study 1 and Study 2 across the 10 categories of agents. There was a strong and significant correlation between the two studies on racial/ethnic recognition accuracy, r = .86, p = .002; gender ratings, r = .99, $p < .001$; human-like ratings, r = .78, p = .008; and likability ratings, r = .86. p = .001. We conclude that the pattern of responding for the 10 agent types on each of the four survey items was consistent between the two studies.

Finally, we also found that the ratings of human-likeness and ratings of likability were significantly correlated with each other in Study 1 (r = .90, $p < .001$), Study 2 (r = .85, $p < .001$), and Study 1 and 2 (r = .78, $p < .001$). Moreover, a simple linear regression model also showed that there was a predictive relationship between ratings of human-likeness and likability. To be more specific, in Study 1, with the fitness regression model: Likability = −.27 + .97*Human-likeness, the overall regression was statistically significant, R^2 = .81, F(1, 35) = 151.06, $p < .001$. It was found that human-likeness significantly predicted likability ratings (β = .90), showing that higher ratings of human-likeness of virtual agents predicted higher likability scores. The same pattern was also found in Study 2, with the fitness regression model: Likability = −.59 + .98*Human-likeness, F(1, 34) = 85.64, $p < .001$; R^2 = .72, β = .85, and the combined data of Study 1 and 2, with the fitness regression model: Likability = −.27 + .97*Human-likeness, F(1, 65) = 98.06, $p < .001$; R^2 = .60, β = .78. In conclusion, our findings indicate a significant correlation between the human-likeness of the agents and people's perceived likability and preferences towards specific agents, such that as the agent becomes more human-like, individuals are more inclined to rate it as more likable.

Discussion

Empirical Contributions

This study produced six key findings. First, people are more accurate in perceiving Black and White agents than Indian, Hispanic, or Asian agents. This outcome partially supports the Alliance Hypothesis and Media Equation Theory, as it reveals successful identification and categorization of Black and White agents. It suggests that regardless of whether the virtual agent had a realistic or cartoon-style appearance, although various ethnicities exhibited distinct visual features, the visual cues associated with different racial groups, such as White agents versus Black agents might be more prominent for people, triggering a stronger automatic encoding process.

Second, people are accurate in perceiving male and female agents. This finding aligns with both the Alliance Hypothesis and the Media Equation Theory, demonstrating the accurate distinction between various gender types of agents. Given that gender serves as one of the most salient and robust visual cues for encoding real humans, it is important to note that individuals also can readily categorize the gender of both realistic-style and cartoon-style virtual agents.

Third, people rated all categories of agents as being moderately human-like, although White agents were rated lower than other racial/ethnic categories.

Fourth, people liked White agents less than other racial/ethnic groups and liked male agents less than female agents. The results indicating lower levels of human-likeness and likability for White agents may find support in a previous study by Nowak and Rauh (2005), which suggested that agents perceived as more human-like tend to be viewed as more credible. In the context of this study, the reduced likability of White agents may be related to also being rated lower in human-like appearances. Furthermore, the higher likability of female agents compared to male agents was also consistent with previous research indicating a preference for female agents or virtual instructors, potentially fostering stronger social connections with human learners than the male agents or virtual instructors (ter Stal et al., 2020; Zhao & Mayer, 2023).

Fifth, the patterns of ratings did not differ much between two independent studies employing different onscreen agents, indicating that the results are not specific to a single set of onscreen agents. This result was different from some previous studies. Specifically, although there was still no consensus about either realistic or cartoon-style agent was better, previous studies generally concurred that the degree of anthropomorphism should play a role in their effectiveness and specific impact on human-technology interactions (Nowak & Rauh, 2005; Reuten et al., 2018; Sträfling et al., 2010). This difference in our findings implies that, within a certain range of anthropomorphism, the discernible disparities in how individuals perceive these agents might not be as pronounced. Consequently, both types of agents presented in this study hold the potential for positive effects and could viably serve as virtual instructors in multimedia learning environments.

Lastly, the supplemental analysis provided additional evidence, revealing a strong correlation between the human-likeness of the agents and likability ratings. More precisely, the agents with higher human-likeness ratings were found to be significantly associated with higher likability ratings. While this result contrasts with the findings from the prior study by Reuten et al. (2018), where robots rated as more human-like elicited weaker pupil dilations and difficulties in recognizing their emotional expressions, it aligns with Nowak and Rauh's study (2005), suggesting that agents perceived as more human-like tend to be considered more credible. The discrepancy between our findings and the study by Reuten et al. (2018) may stem from differences in the level of anthropomorphism. The robots in the Reuten et al. (2018) study, with their higher degree of anthropomorphism, closely resembled real humans, and with their only shown robots' face rather than full body, potentially increasing the likelihood of triggering the uncanny valley effect (Mori, 1970; Mori et al., 2012). Hence, one potential explanation for this disparity could be attributed to the degree of resemblance between the virtual agents in our study and actual human appearance. Unlike Reuten's study, our virtual agents displayed full bodies rather than just faces and did not closely resemble actual human appearance like photographs to the same extreme degree as seen in Reuten's study. This distinction may have shielded our participants from the

unsettling sensations often associated with the uncanny valley effect. It suggests that a moderate level of anthropomorphism and fully embodied agents might be two contributing factors in circumventing the uncanny valley effect. Consequently, this might account for our finding of a positive correlation between human-likeness and the likability ratings of the agents.

Theoretical Implications

Our results support the Media Equation Theory in that people can recognize and relate to on-screen agents differently based on the agent's race/ethnicity and gender. In short, people have affective responses to onscreen agents based on their eracial/ethnic and gender characteristics in the same way as they do for real humans. This is consistent with the Media Equation hypothesis that people treat technology-based beings like real humans.

Our results are also partially consistent the Alliance Hypothesis and the Stereotype Content Model, which suggests that race (though not necessarily ethnicity types) and gender are highly salient social cues that elicit stereotypes concerning likability—even when the characters are animated agents.

Practical Implications

These findings suggest that when instructional designers want to produce animated pedagogical agents that will be liked, they might want to make the agents more human-like, avoid While male agents and instead focus on female agents from other racial/ethnic categories. Alternatively, results suggest that the designers of 3D pedagogical agents should make modifications to increase the likability of White male agents. When it is important that students recognize the race/ethnicity of animated pedagogical agents, more work is needed in providing addition cues to increase the recognition of Asians, Hispanics, and Indians whereas students already easily identify Black and White agents. Further work is needed to determine whether students learn better from onscreen agents that they like more.

Overall, the findings of this study represent a crucial initial step in highlighting the feasibility of integrating diverse onscreen virtual agents into multimedia education platforms. These virtual agents can be easily tailored to align with students' backgrounds and preferences, and the study has shown that students can accurately discern racial and gender types while evaluating agent likability. Thus, using various virtual agents as instructors seems to be a promising potential solution to circumvent the limitations tied to human instructors' identities in multimedia lessons and to cultivate more inclusive and representative educational environments, thereby enhancing the learning experiences of students from diverse backgrounds.

Limitations and Future Directions

Due to inadequate numbers of participants in each racial/ethnic category and each gender type, this study did not examine the Matching Hypothesis, which posits that people tend to be attracted to and establish relationships with others who share similarities in a range of characteristics, encompassing demographics, personality, attitudes, values, and physical appearance (Berscheid & Reis, 1998; Kalick & Hamilton, 1996; Murstein, 1980). However, the results might be different with matched participant-agent demographics (i.e., gender and racial/ethnic types). For instance, the gender difference in perception of emotions has been revealed by a previous neural imaging study (Derntl et al., 2010), with females showing stronger activation in emotion-related areas, particularly the amygdala, during empathy tasks. These results indicated potential variations in male and female participants' reactions to the same virtual agents. Therefore, given the predominantly female sample in the present study, future research should recruit more male participants to match gender and explore potential agent preference disparities. Additionally, preferences for various agent types may differ among participants of diverse racial/ethnic categories or gender types, aligning with the Matching Hypothesis. In particular, individuals may pay more attention to and have a stronger preference for agents who share their own ethnic or racial background. In conclusion, future work is needed to determine whether people have higher recognition accuracy scores and higher human-likeness and likability ratings for agents that share their racial/ethnic and gender categories.

Another limitation to consider is the potential sensitivity of the question regarding the likability of the agents. The dichotomy of "likable" versus "unlikable" may oversimplify participants' perceptions and potentially lead to less candid responses. To address this limitation and obtain more nuanced insights, future research could employ more subtle and finely tuned questioning strategies that encourage participants to provide more sincere and detailed assessments of their feelings and preferences toward the agents.

Moreover, while the study's outcomes align with the Media Equation theory and the Alliance Hypothesis when extended to animated agents, it's important to acknowledge that likability ratings may be influenced by additional factors, including individual personal preferences. Given that likability can have different interpretations, such as likability as a friend or attractiveness as a potential partner, participants might understand this item with different meanings, which might affect our results. Considering our research goal of assessing the potential use of virtual agents as online instructors, it is also necessary to incorporate additional questions that explore whether participants prefer specific types of agents to serve as their instructors. Therefore, a next research step is to expand the approach to questioning to yield a richer understanding of participants' attitudes and preferences, contributing to a more comprehensive evaluation of the agents' effectiveness and suitability in an instructional role.

In addition, this study focused solely on the visual appearance of onscreen agents. However, based on the previous research on Media Equation Theory, even subtle cues,

such as voices, can effectively influence participants' perceptions and reactions to technology, even in the absence of embodied agents. Therefore, future work is needed to examine the role of other social cues such as voice.

Moreover, although it is encouraging that we obtained the same pattern of results across two different sets of animated agents, more work is needed to determine whether the results can be replicated with other different sets of agents. Besides, although participants were able to perceive Black and White agents quite accurately, more work is needed to determine which cues can be added to improve identification of other racial/ethnic categories.

Finally, participants based their responses on very short video clips of each agent. Such brief exposure may not offer participants sufficient stimulus to form connections with these agents and evaluate their likability. Furthermore, given the brevity of the video clips in comparison to a full online lecture, it remains uncertain whether these agents would effectively serve as virtual instructors. Therefore, future work is needed to determine whether responses change when participants have greater exposure to an agent. Additionally, in forthcoming studies, it would be valuable to investigate whether the racial/ethnic and gender categorizations of animated pedagogical agents influence student learning processes and outcomes, particularly in a video lecture context with a virtual instructor. Such investigations could shed light on whether adding various types of virtual instructors in the online lessons can indeed effectively cultivate a more inclusive and representative learning environment and positively enhance the learning experiences and performance of students from diverse backgrounds.

Conclusion

This study demonstrates how well people are able to recognize the race/ethnicity and gender of onscreen agents, how strongly they see each type of agent as human-like, and how much they like each type of agent. The findings highlighted several key points: participants were able to identify the racial and gender types of agents. In addition, although all the agents were rated as moderately human-like, White agents were rated to be least human-like and likable. Male agents were rated as less likable than female agents as well. These results, consistent across two independent studies, supported both the Media Equation Theory and Alliance Hypothesis. This study serves as a critical initial step in highlighting the potential of employing diverse onscreen virtual agents in multimedia education platforms. By modifying and using animated virtual agents, we can overcome the constraints tied to human instructors' inherent identities, paving the way for a more inclusive and representative learning environment for students from diverse backgrounds.

Author Contributions

Fangzheng Zhao and Richard E. Mayer both contributed to developing the design of the study, interpreting the results, and writing the manuscript. Fangzheng Zhao took responsibility for creating the materials, running the participants, and tabulating the data.

Nicoletta Adamo-Villani, Christos Mousas, Minsoo Choi, Luchcha Lam, Magzhan Mukanova, and Klay Hauser contributed to developing a character dataset consisting of 37 realistic-style and 36 cartoonish-style 3D virtual agents, created by Reallusion's Character Creator, and created 3-s video clip for each type of 3D virtual agent.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was supported by Grant 2201019 from the National Science Foundation.

Ethical Statement

Ethical Approval

The research was approved by the Institutional Review Board (IRB) at the University of California, Santa Barbara, and followed the standards for research involving human subjects of the American Psychological Association.

Informed Consent

Participants provided informed consent.

ORCID iDs

Fangzheng Zhao https://orcid.org/0000-0002-3575-9039
Richard E. Mayer https://orcid.org/0000-0003-4055-6938
Christos Mousas https://orcid.org/0000-0003-0955-7959
Minsoo Choi https://orcid.org/0000-0001-9459-4070

Data Availability Statement

The video clips, surveys, and datasets for the current study are available in the OSF repository: https://osf.io/zy2tm/?view_only=9e6e67ec00434d15aa052be8d66ccbc8. This dataset was used exclusively for this paper are Zhao et al. (2023).

Notes

1. We used the term race/ethnicity category to indicate that our five categories involve both race (e.g., Asian, Black, White) and ethnicity (e.g., Hispanic, Indian).
2. There were also six non-binary agents in Study 1, but these were not included in the present analyses because there were no non-binary agents in Study 2. The mean rating on the femininity/masculinity scale was 4.01 (SD = 1.03), with "4″ labeled as "non-binary", 1 as

"very feminine" and 7 as "very masculine." The mean proportion of accurate identification of non-binary agents was .21 (SD = .14). This indicates that participants were not accurate in identifying onscreen agents intended to appear as non-binary. The mean human-like rating was 4.76 (SD = .25) and the mean likability rating was 4.29 (SD = .30), which are not significantly different from the overall average.

3. The significant differences were the same when we employed Bonferroni-adjusted pairwise tests, with alpha set at .005, except that the accuracy scores were not significantly different among Asian, Indian, and Hispanic agents.
4. When we employed Bonferroni-adjusted pairwise tests, with alpha set at .005, there was only one significant difference of human-likeness ratings between Indian agents and White agents, $p < .001$.
5. The significant differences were the same when we employed Bonferroni-adjusted pairwise tests, with alpha set at .005.

References

Belanche, D., Casaló, L. V., Schepers, J., & Flavián, C. (2021). Examining the effects of robots' physical appearance, warmth, and competence in frontline services: The humanness-value-loyalty model. *Psychology and Marketing*, *38*(12), 2357–2376. https://doi.org/10.1002/mar.21532

Berscheid, E., & Reis, H. T. (1998). Attraction and close relationships. In D. T. Gilbert, S. T. Fiske, & G. Lindzey (Eds.), *The handbook of social psychology* (4th ed., pp. 193–281). McGraw-Hill.

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.

Cosmides, L., Tooby, J., & Kurzban, R. (2003). Perceptions of race. *Trends in Cognitive Sciences*, *7*(4), 173–179. https://doi.org/10.1016/s1364-6613(03)00057-3

Davis, N., Olsen, N., Perry, V. G., Stewart, M. M., & White, T. B. (2023). I'm only human? The role of racial stereotypes, humanness, and satisfaction in transactions with anthropomorphic sales bots. *Journal of the Association for Consumer Research*, *8*(1), 47–58. https://doi.org/10.1086/722703

Derntl, B., Finkelmeyer, A., Eickhoff, S., Kellermann, T., Falkenberg, D. I., Schneider, F., & Habel, U. (2010). Multidimensional assessment of empathic abilities: Neural correlates and gender differences. *Psychoneuroendocrinology*, *35*(1), 67–82. https://doi.org/10.1016/j.psyneuen.2009.10.006

Devine, P. G., Hamilton, D. L. E., & Ostrom, T. M. (1994). (Eds.), *Social cognition: Impact on social psychology*. Academic Press.

Domagk, S. (2008).*Pädagogische agenten in multimedialen lernumgebungen. empirische studien zum einfluss der sympathie auf motivation und lernerfolg [pedagogical agents in multimedia learning environments. Empirical studies on the influence of likeability on motivation and performance]*. Logos-Verlag.

Domagk, S. (2010). Do pedagogical agents facilitate learner motivation and learning outcomes? The role of the appeal of agent's appearance and voice. *Journal of Media Psychology*, *22*(2), 84–97. https://doi.org/10.1027/1864-1105/a000011

Eberhard, M. J. W. (1975). The evolution of social behavior by kin selection. *The Quarterly Review of Biology*, *50*(1), 1–33. https://doi.org/10.1086/408298

Fiske, S. T. (2018). Stereotype content: Warmth and competence endure. *Current Directions in Psychological Science*, *27*(2), 67–73. https://doi.org/10.1177/0963721417738825

Fiske, S. T., Cuddy, A. J., Glick, P., & Xu, J. (2002). A model of (often mixed) stereotype content: Competence and warmth respectively follow from perceived status and competition. *Journal of Personality and Social Psychology*, *82*(6), 878–902. https://doi.org/10.1037/0022-3514.82.6.878

Hewstone, M., Hantzi, A., & Johnston, L. (1991). Social categorization and person memory: The pervasiveness of race as an organizing principle. *European Journal of Social Psychology*, *21*(6), 517–528. https://doi.org/10.1002/ejsp.2420210606

Kalick, S. M., & Hamilton, T. E. (1996). The matching hypothesis re-examined. *Journal of Personality and Social Psychology*, *51*(6), 673–682. https://doi.org/10.1037/0022-3514.51.4.673

Krämer, N. C., Rosenthal-von der Pütten, A. M., & Hoffmann, L. (2015). Social effects of virtual and robot companions. In S. S. Sundar (Ed.), *The handbook of the psychology of communication technology* (pp. 137–159). Wiley Blackwell.

Kurzban, R., Tooby, J., & Cosmides, L. (2001). Can race be erased? Coalitional computation and social categorization. *Proceedings of the National Academy of Sciences of the United States of America*, *98*(26), 15387–15392. https://doi.org/10.1073/pnas.251541498

Lawson, A., Mayer, R. E., Adamo-Villani, N., Benes, B., Lei, X., & Cheng, J. (2021a). The positivity principle: Do positive instructors improve learning from instruction video lectures? *Educational Technology Research & Development: ETR & D*, *69*(6), 3101–3129. https://doi.org/10.1007/s11423-021-10057-w

Lawson, A. P., & Mayer, R. E. (2021). The power of voice to convey emotion in multimedia instructional messages. *International Journal of Artificial Intelligence in Education*, *32*(4), 971–990. https://doi.org/10.1007/s40593-021-00282-y

Lawson, A. P., Mayer, R. E., Adamo-Villani, N., Benes, B., Lei, X., & Cheng, J. (2021b). Do learners recognize and relate to the emotions displayed by virtual instructors? *International Journal of Artificial Intelligence in Education*, *31*(1), 134–153. https://doi.org/10.1007/s40593-021-00238-2

Lawson, A. P., Mayer, R. E., Adamo-Villani, N., Benes, B., Lei, X., & Cheng, J. (2021c). Recognizing the emotional state of human and virtual instructors. *Computers in Human Behavior*, *114*(4), 106554. https://doi.org/10.1016/j.chb.2020.106554

Messick, D. M., & Mackie, D. M. (1989). Intergroup relations. *Annual Review of Psychology*, *40*(1), 45–81. https://doi.org/10.1146/annurev.ps.40.020189.000401

Michod, R. E. (1982). The theory of kin selection. *Annual Review of Ecology and Systematics*, *13*(1), 23–55. https://doi.org/10.1146/annurev.es.13.110182.000323

Murstein, B. I. (1980). Mate selection in the 1970s. *Journal of Marriage and Family*, *42*(4), 777–792. https://doi.org/10.2307/351824

Nass, C., Moon, Y., & Green, N. (1997). Are machines gender neutral? Gender-stereotypic responses to computers with voices. *Journal of Applied Social Psychology*, *27*(10), 864–876. https://doi.org/10.1111/j.1559-1816.1997.tb00275.x

Nass, C., & Steuer, J. (1993). Voices, boxes, and sources of messages: Computers and social actors. *Human Communication Research*, *19*(4), 504–527. https://doi.org/10.1111/j.1468-2958.1993.tb00311.x

Nowak, K. L., & Rauh, C. (2005). The influence of the avatar on online perceptions of anthropomorphism, androgyny, credibility, homophily, and attraction. *Journal of Computer Mediated Communication*, *11*(1), 153–178. https://doi.org/10.1111/j.1083-6101.2006.tb00308.x.

Pietraszewski, D. (2016). Priming race: Does the mind inhibit categorization by race at encoding or recall? *Social Psychological and Personality Science*, *7*(1), 85–91. https://doi.org/10.1177/1948550615602934

Pietraszewski, D. (2021). The correct way to test the hypothesis that racial categorization is a byproduct of an evolved alliance-tracking capacity. *Scientific Reports*, *11*(1), 3404. https://doi.org/10.1038/s41598-021-82975-x

Pietraszewski, D. J. (2009). *Erasing race with cooperation: Evidence that race is a consequence of coalitional inferences*. University of California.

Reeves, B., & Nass, C. (1996). *The media equation: How people treat computers, television, and new media like real people*. Cambridge University Press.

Reuten, A., Van Dam, M., & Naber, M. (2018). Pupillary responses to robotic and human emotions: The uncanny valley and media equation confirmed. *Frontiers in Psychology*, *9*(1), 774. https://doi.org/10.3389/fpsyg.2018.00774

Rosenthal-Von Der Pütten, A. M., Schulte, F. P., Eimler, S. C., Sobieraj, S., Hoffmann, L., Maderwald, S., Brand, M., Krämer, N. C., & Krämer, N. C. (2014). Investigations on empathy towards humans and robots using fMRI. *Computers in Human Behavior*, *33*(1), 201–212. https://doi.org/10.1016/j.chb.2014.01.004

Stangor, C., Lynch, L., & Glas, B. (1992). Categorization of individuals on the basis of multiple social features. *Journal of Personality and Social Psychology*, *62*(2), 207–218. https://doi.org/10.1037/0022-3514.62.2.207

Sträfling, N., Fleischer, I., Polzer, C., Leutner, D., & Krämer, N. C. (2010). Teaching learning strategies with a pedagogical agent. The effects of a virtual tutor and its appearance on learning and motivation. *Journal of Media Psychology*, *22*(2), 73–83. https://doi.org/10.1027/1864-1105/a000010

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

Taylor, S. E., Fiske, S. T., Etcoff, N. L., & Ruderman, A. J. (1978). Categorical and contextual bases of person memory and stereotyping. *Journal of Personality and Social Psychology*, *36*(7), 778–793. https://doi.org/10.1037/0022-3514.36.7.778

Zhao, F., & Mayer, R. E. (2023). *Role of emotional tone and gender of computer-generated voices in multimedia lessons* (pp. 1–21). Educational Technology Research and Development.

Zhao, F., Mayer, R. E., Adamo-Villani, N., Mousas, C., Choi, M., Lam, L., Mukanova, M., & Hauser, K. (2023). Using AI to Transform Online Video Lectures into Effective and Inclusive Agent-Based Presentations [dataset]. OSF repository: https://doi.org/10.17605/OSF.IO/ZY2TM

Author Biographies

Fangzheng Zhao is a 4th year PhD Candidate in Dr. Richard Mayer's Lab at University of California, Santa Barbara. Her research focuses on multimedia learning, to be more specific, exploring various strategies to improve the effectiveness of video learning or game-based learning.

Richard E. Mayer is Distinguished Professor of Psychology at the University of California, Santa Barbara. His research interests are in applying the science of learning to education, with current projects on multimedia learning, computer-supported learning, and computer games for learning.

Nicoletta Adamo-Villani is a Professor of Computer Graphics Technology and Purdue University Faculty Scholar. She is an award-winning animator and graphic designer and creator of several 2D and 3D animations that aired on national television. Her area of expertise is in character animation and character design and her research interests focus on the application of 3D animation technology to education, HCI (Human Computer Interaction), and visualization.

Christos Mousas is an Associate Professor in the Department of Computer Graphics Technology and director of the Virtual Reality Lab at Polytechnic Institute, Purdue University. His research revolves around virtual reality, virtual humans, computer graphics & animation, and human-computer interaction. Specifically, he uses computer graphics and computational tools to design virtual reality experiences, and aspects of cognitive and experimental psychology to understand how humans interact in virtual environments.

Minsoo Choi is a Ph.D. student in the Department of Computer Graphics Technology at Purdue University. His research interests include Virtual Reality, Human-Computer Interaction, and Computer Animation. Specifically, in his research, Minsoo focuses on how people interact with virtual agents in virtual environments.

Luchcha Lam earned her master's degree in computer graphics technology at Purdue University in May 2023. She was trained as both a programmer and an artist with experience in the gaming and animation industries. Her expertise is in computer graphics, machine learning, and game design.

Magzhan Mukanova is a master's student in Computer Graphics Technology in 3D Animation at Purdue University with the background in Engineering and Technology (BS).

Klay Hauser is from West Lafayette Indiana. He studied at Purdue University for his Bachelor of Science in Animation. He is currently studying at Purdue University for his Masters in Computer Graphics Technology.