Effects of Rhetorical Strategies and Skin Tones on Agent Persuasiveness in Assisted Decision-Making

Amama Mahmood
amahmo11@jhu.edu
Johns Hopkins University
Baltimore, Maryland, USA

Chien-Ming Huang
chienming.huang@jhu.edu
Johns Hopkins University
Baltimore, Maryland, USA

ABSTRACT

Appearance and linguistic cues may influence how both people and Intelligent Virtual Agents (IVAs) are perceived and evaluated by others; appearance (e.g., skin tone) has been linked to various implicit biases such as agreeing more with stereotypical attractive faces, while particular linguistic cues may effectively increase persuasiveness. In this paper, we report an online study (N=59) evaluating how strategic linguistic cues (expertise: high vs. low) may shape the implicit disadvantages associated with ethnic stereotypes (skin tone: dark vs. light). We found that a virtual agent with a high level of expertise was considered more persuasive, dominant, intelligent, and likeable regardless of their skin tone, and that participants complied more with IVAs with a darker skin tone. Our results suggest that the design of IVAs requires the deliberate considerations of factors such as appearance and linguistic behaviors in order to achieve intended outcomes.

CCS CONCEPTS

• **Human-centered computing** → **Empirical studies in HCI**; • **Computing Methodologies** → *Artificial intelligence.*

KEYWORDS

Human-agent interaction, embodied virtual agent, persuasiveness, skin-tone biases

ACM Reference Format:
Amama Mahmood and Chien-Ming Huang. 2022. Effects of Rhetorical Strategies and Skin Tones on Agent Persuasiveness in Assisted Decision-Making. In *Proceedings of IVA '22: International Conference on Intelligent Virtual Agents (IVA '22).* ACM, New York, NY, USA, 8 pages. https://doi.org/10.1145/nnnnnnn.nnnnnnn

1 INTRODUCTION

Intelligent Virtual Agents (IVAs) are envisioned to assist people in making important decisions by providing suggestions or expert advice in personal settings as assistants, trainers, health coaches, and recommenders in online services, as well as in professional settings as receptionists, career counselors, healthcare professionals, and financial advisors. For example, the recent launch of Microsoft Mesh and Metaverse has given users a social virtual reality (VR) platform to explore and create content, which calls for the use of IVAs for assisting users in making important decisions while navigating these virtual spaces.

While the *embodiment* of intelligent agents significantly improves users' perception of social presence [37], motivation [6], rapport [44], and trust [7, 44], it brings its own set of design challenges. For instance, implicit skin-tone biases commonly found against people of color have been observed in human-agent and human-robot interactions [4, 42]. The "attractiveness" of a virtual agent has also been found to influence people's perceptions of the persuasiveness of the agent [28]. Moreover, the modulation of gender presentation triggers gender stereotypes and sets gender-based expectations [17, 35].

As IVAs continue to emerge as digital aids for assisted decision making, it is imperative to ensure that these technologies do not replicate or reinforce existing societal biases. Though increasing evidence suggests that appearance of embodied agents alone can greatly affect how they are perceived, it is alarming to see that these agents are predominantly represented as white, in color, ethnicity or both [33, 48]. This unequal representation reinforces skin-tone biases similar to what we see in human-human interactions; it was recently reported that intelligent, professional, and powerful agents are essentially "White" [13]. As IVAs continue to mature and become more available for delivering expert advice and engaging in co-decision making, they will interact with more diverse users. The impact of implicit biases emerged in human-agent interactions could be consequential, especially in critical domains such as healthcare. To reduce implicit biases stemming from the skin-tone of virtual agents, previous research has investigated leveraging empathy games to elicit compassion for underrepresented populations [22], and using avatar embodiment representing black skin-tone [2]. While these explorations have offered valuable insights into how out-of-interaction alteration may help reduce skin-tone biases, it is unclear how an IVA may actively combat implicit bias through its behavior during the interactions.

The success of IVAs—whether in motivating users to adhere to an exercise plan or in encouraging them to open up in therapy—will depend on their abilities to persuade people. Research in Human-Robot Interaction (HRI) has identified various rhetorical linguistic cues capable of increasing people's perception of a robot's expertise [1] and illustrated that an expert robot is perceived more competent and able to convince people into compliance more than a non-expert one [24]. In this work, we explore how linguistic cues of expertise may be applied to IVAs and whether agents with higher level of expertise can overcome possible skin-tone biases. To study this, we conducted an online study where a virtual agent varying

IVA '22, September 06–September 09, 2022, Faro, Portugal

ACM ISBN 978-x-xxxx-xxxx-x/YY/MM…$15.00
https://doi.org/10.1145/nnnnnnn.nnnnnnn

in expertise (*expert vs. novice*) and skin color (*dark vs. light*) tried to persuade participants in a desert survival task. Our results indicate a clear effect of rhetorical linguistic cues on people's compliance and their perceived persuasiveness of the agent. However, in contradiction to prior work, we see a slight preference towards agents with a darker skin tone. Next, we review relevant prior research that motivates this work.

2 BACKGROUND AND RELATED WORK

2.1 Skin Tone, Race and Ethnicity

Race and ethnicity are often interchangeably used, however, according to anthropologists there is only one human race [8]. Therefore, in this paper, we use the term "ethnicity" instead of "race". Ethnicity is a complex social construct that intersects with a range of cultural factors and biological characteristics. The complexity of ethnicity has also been found in human-agent interaction (e.g., [50]). In many prior works involving embodied virtual agents and social robots, experimental manipulations of ethnicity were just simplified through the manipulations of skin tone (e.g., [2, 4, 5, 26, 42]). In this present work, we acknowledge the multi-faceted nature of ethnicity and culture and focus on skin tone without explicitly imposing an ethnic identity to an embodied virtual agent. We believe that this present investigation is particularly timely provided the recent rise in awareness of colorism and the needs for a critical reflection on ethnicity in the HCI community [38].

2.2 Implicit Biases Towards Embodied Agents

Implicit bias commonly found against people of color has also been observed in the domains of human-agent and human-robot interactions. For instance, a virtual human patient with a light skin tone (white) elicited more empathetic behavior from medical students than the one with a darker skin tone (black) [42]. Previous research has also found biases in how people reacted to colored robots [4]. However, the implicit bias can potentially be leveraged for a positive outcome. For example, expert agents (realistic in appearance) in non-traditional representations, namely Black instead of White, caused higher transfer of learning and enhanced focus and concentration for students despite that the agents used the identical messages [5]; the authors speculated that the Black agents were perceived to be "novel" thus warranting more attention.

2.3 Strategies to Reduce Implicit Biases

Various strategies such as perspective taking to imagine oneself as part of the stigmatized group and increasing exposure to outgroup result in significant reductions in implicit racial biases and enhanced personal awareness of these biases [16]. Strategic interventions of presenting users with counter-stereotypical exemplars, e.g., portraying a black man as a rescuer while a white man assaulted the participant in an evocative story, have shown to be effective in combating implicit biases [30]. Similar strategies to reduce biases associated with gender and ethnicity have been explored in HCI as well. One of these strategies is using empathy games that employ graphic stories in hope to elicit compassion and comprehension for underrepresented populations by exposing the players to experiences of others. For instance, engaging in *Fair Play*, an interactive video game designed to elicit empathy for black graduate students from non-black populations in predominantly white universities, resulted in reduction of implicit racial biases [22].

In virtual environments, embodied identification have been found useful in reducing implicit societal biases [26]. For example, white female participants, who were represented by a black avatar in VR and showed strong identification with the avatar, exhibited sustained reduction of implicit biases [2]. Embodied identification has shown to be useful in reducing biases more for the group which is biased against the most by the most privileged group e.g., black female personas are biased against the most by white male participants [26]. In this work, we sought to explore the use of linguistic cues of expertise (rhetorical speech) as an active strategy to minimize possible implicit biases in human-agent interactions.

2.4 Designing Persuasive Embodied Agents

Persuasion is defined as a process in which a person attempts to influence or control another person's decision making, opinion, or behavior [18]. In human-human interactions, non-verbal cues, such as gestures, body movements, facial expressions, and eye contact, play an key role in shaping credibility and hence persuasiveness [9]. In addition to non-verbal cues, vocal cues such as tempo, pitch, and loudness influence how persuasive a person is perceived [49]. Research in HRI also explored how robots might deliver information effectively to convince users into compliance [32, 43]. Prior works have indicated various factors that may influence persuasiveness in robots including height [41], facial characteristics [19], bodily cues [15], and gender [21, 46]. Human-inspired expert linguistic cues were also explored to increase the perceptions of expertise for robots [1]. Furthermore, research has indicated that people perceive an expert robot to be more competent and would comply more to it compared to a non-expert robot [24]. As informed by these works on designing persuasive robots, this current study explores how some of the effective strategies may be translated to IVAs.

3 METHODS

This section details our study exploring how an avatar's skin tone and rhetorical speech affect people's interactions with it.

3.1 Hypotheses

We formulated the following hypotheses based on previous persuasive communication and implicit bias research:

- **Hypothesis 1.** Participants will comply with an expert agent's reasoning more than they will with a novice agent; in particular, an expert agent employing rhetorical strategies will be perceived as more persuasive, intelligent, and likeable than a novice, as informed by prior works [1, 23, 24].
- **Hypothesis 2.** Agents with light (white) skin tones will be perceived more positively than agents with dark (black) skin tones, regardless of expertise, as informed by prior works (e.g., [36, 42]).
- **Hypothesis 3.** The expert agent's speech strategy will decrease the reflection of implicit biases in participants' behavior; in particular, we predict that the expert's quality of speech will reduce implicit biases against black or in favor of white agents when presented in non-traditional representations [5] or as counter-stereotypical examples i.e., black agents presented as experts and white agents as novices. [30].

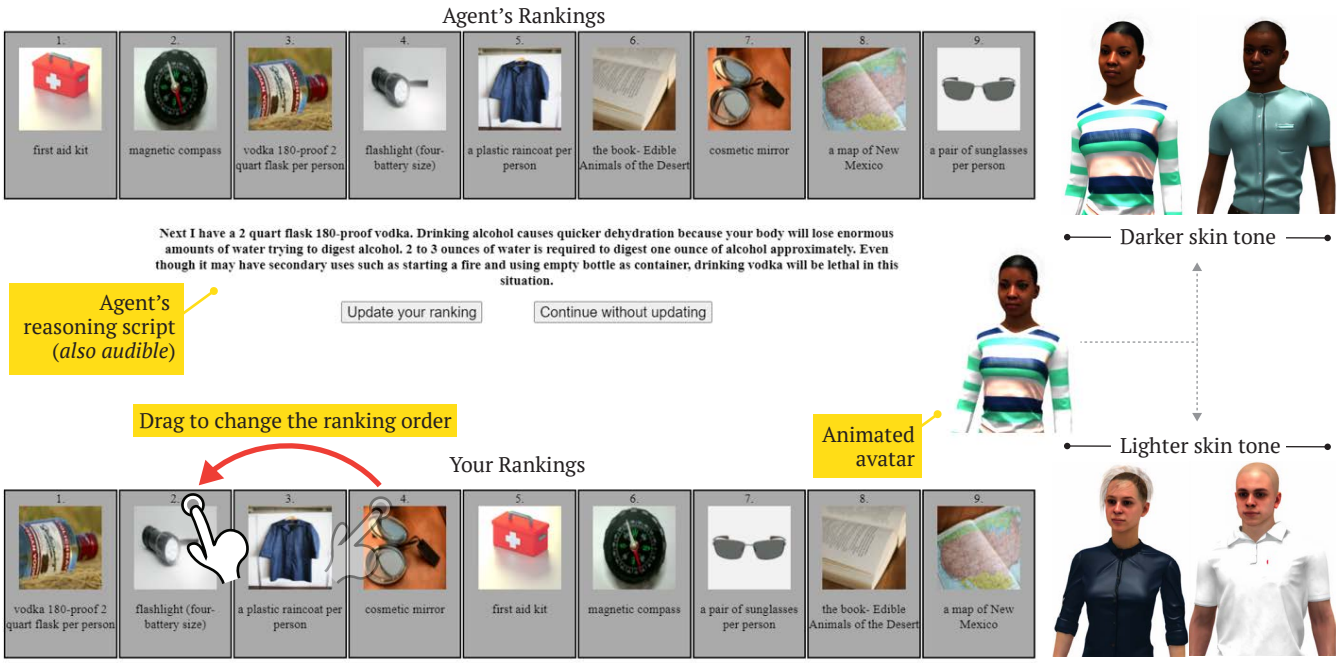

Figure 1: The web interface for our experiment. The AI agent provides its reasoning to persuade a participant to change their ranking. The participant could choose to rearrange their ranking by clicking on the "Update your ranking" button, which would allow them to drag and drop the items to update their ordering. Four different avatars were used in our study.

3.2 Experimental Design and Task

We conducted a between-subjects study consisting of four experimental conditions, following a 2x2 factorial design with two independent variables: skin tone (*black/dark* vs. *white/light*) and linguistic cues of expertise (*expert* vs. *novice*) exhibited by the IVA. We contextualized our investigation using an updated version of the Desert Survival Situation [29], as informed by prior works that study team communication behavior [40, 41].

We conducted our study online via a web application; participants were given a detailed description of a plane crash in the New Mexico desert and then asked to rank nine items (presented in a randomized order to each participant) in order of importance for their own hypothetical survival (Fig. 1). After finalizing their initial rankings of the items, participants were instructed to interact with an embodied virtual agent (Fig. 1). A ranking list for the agent was automatically generated; the agent's ranking was consistently different from that of the participant. We employed the same algorithm for generating the agent's ranking as used in prior studies [29, 40, 41] (supplementary materials B[1]). During the subsequent interaction, the agent provided reasoning for its rankings, item by item; the agent listed several facts about the current item to convince the participant to change their ranking of it. Participants had the opportunity to update their rankings or continue without doing so. The agent then moved on to the next item. If the agent and participant agreed on the placement of an item, the agent skipped its explanation and proceeded to the next item after saying: "Glad we agree on some items on the list."

3.3 Agent Design and Assignment

Our agent was embodied as a virtual avatar and communicated with participants via both voice and text. We designed four avatar options and their behaviors in Mixamo[2]. The agents were not identical in facial features and clothing to offer more generalizability of our results in terms of avatar design. The agent had two non-verbal behaviors, *idle* and *talking*, both of which were equivalent across experimental conditions. The talking behavior activated when the agent explained its reasoning for a particular item's ranking; otherwise, the agent performed the idle behavior.

We did not manipulate gender in this study. However, various effects of gender and gender differences have been reported in prior HCI research; for instance, when a participant is presented with an agent of the same gender, they exhibit increased trust [17], lower negative emotions [19], and higher change in attitude [21]. Since the effect of gender on persuasiveness is not the focus of our study, we decided to match the agent's gender to that of each participant. Based on their self-identified gender, our system assigned an avatar that matched the participant's gender; if a participant identified as "non-binary" or "other," they were randomly assigned a male- or female-presenting avatar. To generate our agent's speech, we used Microsoft's Text to Speech service to convert the explanatory text passages to audio files; we used Microsoft voices "David" and "Zira" with default parameters for male and female voices, respectively.

3.4 Manipulations and Conditions

3.4.1 Skin Tone. While skin color and its connotations are much more nuanced in reality, in this study we focused on two skin tones:

[1]https://intuitivecomputing.jhu.edu/publications/2022-iva-mahmood-supp.pdf

[2]https://www.mixamo.com/

black/dark and white/light. Once a gender was assigned to the agent, our system randomly assigned the agent a skin tone: black or white. For instance, a female participant might interact with either white or black female agent.

3.4.2 Linguistic Cues of Expertise. The agent followed a scripted monologue (see supplementary materials C.2[1]) that included the advantages and disadvantages of each of the nine items. If the participant ranked an item as less important on their list than the agent had, the agent highlighted the advantages of that item; in the opposite case, the agent highlighted the item's disadvantages to convince the participant that it should be ranked as less important. We manipulated the linguistic cues of expertise to create two types of speech: *expert* and *novice*. We built upon previous work designing rhetorical robots with linguistic cues of expertise; for our study, we replicated, modified, and added to a previously explored model of expert speech [1], using a combination of the following aspects to create the two types of speech for our study:

- **Level of practical knowledge.** Experts possess more specific knowledge and are perceived as more knowledgeable in their field than the average person [23]. We, therefore, included fewer details about the survival items in the novice's monologue and more details and reasoning in the expert's monologue.
- **You-perspective.** You-perspective, also known as "you-attitude," or the second-person point of view, makes an audience feel more valued; in business communication, you-perspective is most frequently used by professional negotiators (rather than mere aspiring ones) [39, 45]. For our novice's speech, we used "people" in place of "you" to alienate the audience.
- **Fluency.** Long pauses are perceived as disruptive and hesitation is seen as an indication of inexperience [11, 12]; therefore, we programmed our novice speech with pauses that were five times longer [11] than those in our expert's speech. A standard pause was represented by one full stop (".") in text, whereas a longer pause was represented by five full stops ("....."); in our text-to-speech module, the five full stops translated to a pause that was five times longer than a normal pause between sentences. Furthermore, we used sentence breakers and phrases—such as "I think," "I suppose," "um," "oh," and "that's it"—to depict uncertainty and lack of fluency. In our novice agent's speech, we used three breakers or pauses, randomly chosen, in each statement.
- **Organization.** A logical progression of information with sufficient detail and good flow is representative of an expert speaker [1, 20, 34]; thus, we ensured a proper flow of useful information without interruptions for our expert speech. On the other hand, we included sentence breakers (explained above) and poor sentence construction in our novice speech.

To see the difference between expert and novice speech for all nine survival items, please refer to supplementary materials C.2[1].

3.5 Measures

We used a range of objective and subjective metrics to measure task compliance, persuasiveness, and other perceptions of the agent.

3.5.1 Verification of avatar appearance. We used two checks to verify if the agents' skin tone and gender were perceived as intended (supplementary materials A[1]).

3.5.2 Task compliance. We measured participant compliance via the difference between the agent's rankings and the participants' final rankings; note that the agent's rankings did not change throughout the interaction. Since the agent's rankings were designed to be consistently different from all participants' initial rankings, the difference between those two rankings was consistent across trials. Therefore, the difference between the agent's ranking and a participant's final ranking is a direct measure of how much influence the agent had on that participant.

- **Spearman's** ρ (Range: 0–1). We computed Spearman's ρ (a rank-order correlation) to assess the similarity between the agent's and a participant's rankings. A smaller ρ implied that the participant was less persuaded by the agent to change their initial ranking.
- **Cumulative changes.** This metric captures the degree to which the participant changed the rank of the item under consideration. The measure sums over all nine items, $\sum_{i=1}^{9}(|i - prev\_rank| - |i - current\_rank|)$, where i is the index of the current item on the agent's ranking list $agent\_list$; $prev\_rank$ is the rank of the current item ($agent\_list(i)$) in the participant's list before the the agent presented its reasoning; and $current\_rank$ is the rank of the current item on the participant's list after the agent's reasoning. A positive value suggests that the participant moved their item closer to the agent's suggested rank.

3.5.3 Subjective evaluation.

- **Persuasiveness** (Three items; Cronbach's $\alpha = .80$). We adapted a persuasiveness scale from a prior study [15] to measure the persuasive ability of our agent; the questions we included were: The agent is 1) dissuasive – persuasive and 2) unhelpful – helpful, and 3) The agent's content is: irrelevant – relevant. We used a five-point rating scale.
- **Interpersonal Dominance Scale** (Five items; Cronbach's $\alpha = .76$). We adapted and modified the Interpersonal Dominance Scale [10] to measure the dominance of the agent by asking participants to rate five statements: 1) *The agent rarely influenced me*; 2) *The agent often stopped to think about what to say next*; 3) *The agent often had trouble thinking of things to talk about*; 4) *The agent was not very smooth verbally*; and 5) *The agent was usually successful in persuading me to change my ranking*. We used a five-point Likert scale with 1 being "Strongly disagree" and 5 being "Strongly agree."
- **Perceived Intelligence** (Three items; Cronbach's $\alpha = .93$). We used the Godspeed questionnaire [3] to measure the agent's perceived intelligence. On a scale of 1 to 5, participants rated their impression of the agent: 1) Incompetent – Competent, 2) Ignorant – Knowledgeable, and 3) Foolish – Intelligent.
- **Likeability** (Three items; Cronbach's $\alpha = .70$). We also employed the Godspeed questionnaire [3] to gauge the agent's likeability. On a scale of 1 to 5, participants rated their impression of the agent: 1) Unfriendly – Friendly, 2) Unpleasant – Pleasant, and 3) Awful – Nice.

3.6 Procedure

Upon consenting to participate in this study, participants filled a demographics survey and completed the experimental task. Afterwards, participants filled out a questionnaire regarding their

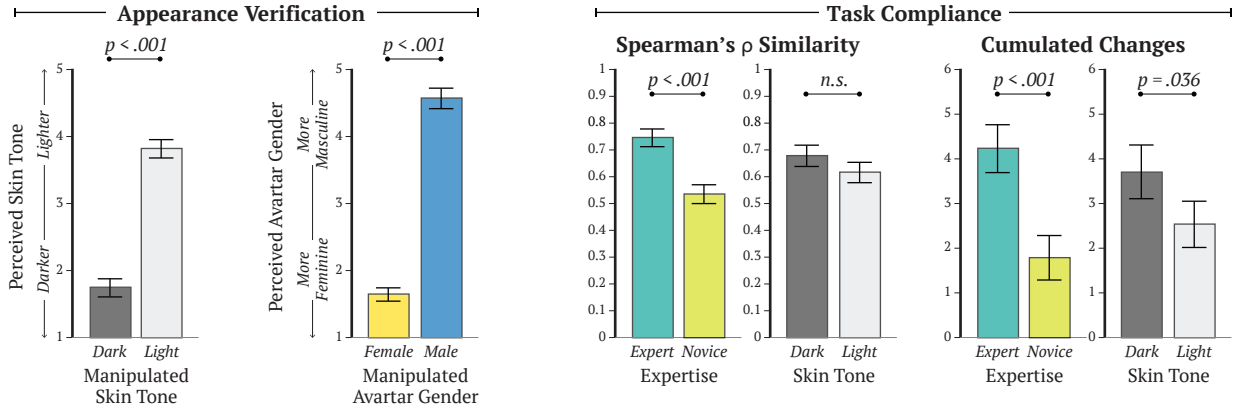

Figure 2: Results on verification of agent's skin tone and gender (left), and task compliance metrics (right). A Welch's t-test was conducted to verify manipulation of skin tone and gender of agents. A two-way ANOVA was conducted to discover effects of agent's skin tone (*light/white vs. dark/black*) and level of expertise (*expert vs. novice*) on Spearman's ρ and cumulated changes. Error bars represent standard error (SE).

perceptions of the agent. They were randomly assigned to one of the four conditions, although we ensured that we had at least 10 participants per condition and that within each of those conditions, at least five participants interacted with male agent and five interacted with female agent. The study took approximately 20 minutes to complete and participants were compensated with a $5 gift card. The study was approved by our institutional review board (IRB).

3.7 Participants

Fifty-nine participants (36 female, 23 male) were recruited for this online study via convenience sampling. The participants were aged 18 to 59 years ($M = 23.51, SD = 7.79$), currently living in the United States of America, and had a variety of educational and professional backgrounds, such as computer science, robotics, healthcare, life sciences, and education. The participants self-identified as Asian ($n = 34$), African American ($n = 7$), Caucasian ($n = 11$), and Hispanic, Latino, or Spanish origin ($n = 7$). A total of 12 participants interacted with black expert agent (7 female, 5 male), 19 with the white expert (14 female, 5 male), 15 with the black novice (8 female, 7 male), and 13 with the white novice (7 female, 6 male).

4 RESULTS

The results reported below, unless specified otherwise, were based on two-way analysis of variance (ANOVA) where *expertise* and *skin tone* were set as fixed effects. The assumptions of two-way ANOVA are no significant outliers, approximate normal distribution, and homogeneity of variances. Our data had no significant outliers. Shapiro-Wilk test validated approximate normal distribution assumption for dependent variables. Moreover, Levene's test revealed that assumption of homogeneity of variances holds. All post-hoc pairwise comparisons were conducted using Tukey's HSD test. Figs. 2 and 3 summarize our main findings.

4.1 Verification of Avatar Appearance

We used Welch's t-tests, assuming unequal variances, to verify the agents' appearances—skin tone and gender—were perceived by the participants as intended (Fig. 2, left).

4.1.1 Skin tone. The 32 participants who interacted with the white agent ($M = 3.82, SD = .78$) compared to the 27 participants who interacted with the black agent ($M = 1.74, SD = .71$) rated the agent's skin tone to be significantly lighter, $t(56.62) = 10.65, p < .001$, indicating that our manipulation of skin tone was adequate.

4.1.2 Gender. The 36 participants who interacted with the female agent ($M = 1.64, SD = .59$) compared to the 23 participants who interacted with the male agent ($M = 4.56, SD = .73$) indicated the agent's gender as more female than male significantly, $t(40.08) = 16.16, p < .001$.

4.2 Task Compliance

4.2.1 Spearman's ρ. A two-way ANOVA test yielded a significant main effect of the agent's expertise (novice: $M = 0.53, SD = 0.19$ vs. expert: $M = 0.74, SD = 0.19$), $F(1, 55) = 21.25, p < .001, \eta_p^2 = .279$, on Spearman's ρ. We found no main effect of the agent's skin tone (dark: $M = 0.68, SD = 0.21$ vs. light: $M = 0.62, SD = 0.21$), $F(1, 55) = 3.94, p = .052, \eta_p^2 = .067$, on this similarity measure. No significant interaction effect between expertise and skin tone was found, $F(1, 55) = .281, p = .598, \eta_p^2 = .005$.

4.2.2 Cumulated changes. A two-way ANOVA test yielded significant main effects of expertise (novice: $M = 1.79, SD = 2.63$ vs. expert: $M = 4.23, SD = 2.97$), $F(1, 55) = 13.71, p < .001, \eta_p^2 = .199$, and skin tone (dark: $M = 3.79, SD = 3.12$ vs. light: $M = 2.53, SD = 2.93$), $F(1, 55) = 4.63, p = .036, \eta_p^2 = .078$, on cumulated changes. No significant interaction effect was found, $F(1, 55) = .11, p = .741, \eta_p^2 = .002$.

4.3 Subjective Evaluation

4.3.1 Persuasiveness. A two-way ANOVA test yielded a significant main effect of the agent's expertise (novice: $M = 2.72, SD = 0.77$ vs. expert: $M = 3.86, SD = 0.62$), $F(1, 55) = 38.44, p < .001, \eta_p^2 = .411$, on the perceived persuasiveness of the agent. We found no significant main effect of the agent's skin tone (dark: $M = 3.31, SD = 0.90$ vs. light: $M = 3.33, SD = 0.90$), $F(1, 55) = 0.65, p = .424, \eta_p^2 = .012$

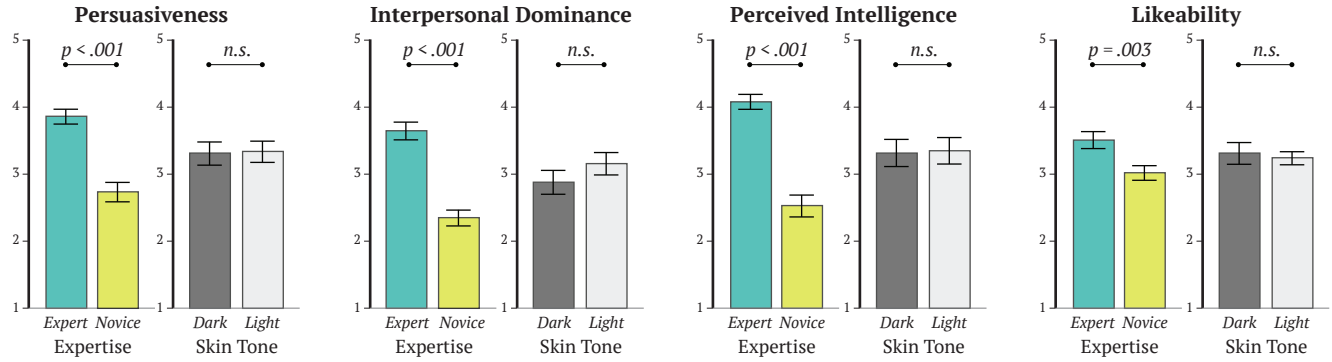

Figure 3: Results on subjective measures: persuasiveness, interpersonal dominance, perceived intelligence, and likeability. Two-way ANOVAs were conducted to discover effects of agent's skin tone (*light/white vs. dark/black*) and level of expertise (*expert vs. novice*) on subjective measures. Error bars represent standard error (SE).

on persuasiveness. We also did not observe a significant interaction effect, $F(1, 55) = 0.10, p = .759, \eta_p^2 = .002$.

4.3.2 Interpersonal Dominance. A two-way ANOVA test yielded a significant main effect of the agent's expertise (novice: $M = 2.36, SD = 0.63$ vs. expert: $M = 3.65, SD = 0.73$), $F(1, 55) = 48.04, p < .001, \eta_p^2 = .466$, on perceived agent dominance. The test found no significant main effect of the agent's skin tone (dark: $M = 2.89, SD = 0.92$ vs. light: $M = 3.17, SD = 0.95$), $F(1, 55) = 0.21, p = .647, \eta_p^2 = .004$ on perceived dominance. No significant interaction effect was observed, $F(1, 55) = 1.44, p = .235, \eta_p^2 = .026$.

4.3.3 Perceived Intelligence. A two-way ANOVA test yielded a significant main effect of the agent's expertise (novice: $M = 2.52, SD = 0.87$ vs. expert: $M = 4.08, SD = 0.63$), $F(1, 55) = 62.13, p < .001, \eta_p^2 = .530$, on perceived intelligence. However, we did not observe a significant main effect of the agent's skin tone (dark: $M = 3.32, SD = 1.05$ vs. light: $M = 3.35, SD = 1.12$), $F(1, 55) = 1.10, p = .299, \eta_p^2 = .020$ on perceived intelligence, nor did we see a significant interaction effect, $F(1, 55) = 0.96, p = .331, \eta_p^2 = .017$.

4.3.4 Likeability. A two-way ANOVA test revealed a significant main effect of the agent's expertise (novice: $M = 3.01, SD = 0.58$ vs. expert: $M = 3.50, SD = 0.70$), $F(1, 55) = 9.68, p = .003, \eta_p^2 = .150$, on likeability. No significant main effect of the agent's skin tone (dark: $M = 3.31, SD = 0.84$ vs. light: $M = 3.24, SD = 0.54$), $F(1, 55) = 0.68, p = .415, \eta_p^2 = .012$, on likeability was observed. We also did not see a significant interaction effect, $F(1, 55) = 1.57, p = .216, \eta_p^2 = .028$.

5 DISCUSSION

This work examined how *expertise* (expert vs. novice) achieved through the employment of rhetorical linguistic cues and *skin tone* (dark or light) of an embodied IVA may affect participants' compliance with and perceptions of the agent. Below, we discuss our results, their design implications, limitations, and future research.

5.1 Expert Agents are Persuasive

Our hypothesis 1 predicts that the expert agent will be considered more persuasive, intelligent, and likeable. Our results support this hypothesis and show that the rhetorical cues of expertise significantly influenced participants' decisions to update their lists (Fig. 2, right). Participants were more compliant with the reasoning given by the expert agent. The participants in the expert conditions made more changes to their lists to match the ranking of the agent more closely. Moreover, the expert agent was perceived to be more persuasive, intelligent, and likeable (Fig. 3). Overall, our results regarding effectiveness of rhetorical strategies on persuasion and perceptions of agent's intelligence and likeability are in line with the findings of the previous studies in HRI [1, 23, 24].

5.2 Preferring Agents with a Darker Skin Tone

As informed by prior research suggesting that virtual agents face skin-tone bias just as humans do [42], we hypothesized an overall greater compliance with and positive impression of white agents in comparison to the black. Our results, however, indicate that the participants did not comply more with the white agents. Furthermore, we did not see any significant preference toward agents with light skin tones, regardless of their expertise levels, for the measures of persuasiveness, dominance, perceived intelligence, and likeability (Fig. 2 (right) and Fig. 3). Instead, we observed a slight preference towards agents with darker skin; for instance, we found that agents with darker skin tones convinced the participants to make more changes (cumulated changes) than the white agents. All in all, our results did not support Hypotheses 2 and 3.

We speculate that these results may be attributed to a number of reasons. First, our avatars may not have adequate fidelity to afford humanlike interactions. In fact, the participants rated the agents to be very machinelike ($M = 1.88, SD = 1.12$) on a scale of 1–5 (1 being very machinelike and 5 being very humanlike). As a result, implicit racial biases in human interactions might not be carried over to our online interaction setting. Second, the recent rise of awareness of bias against people of color may have made participants more conscious about how they evaluated the agents. Furthermore, we suspect that the participants may be displaying "in-group bias" e.g., people with darker skin tone favoring black agents, and using "active overcompensation" for out-group members [47], a performance-oriented behavior, e.g., people with lighter skin tone agreeing more with the black agents to appear warmer than they normally would towards black agents to avoid appearing colorist

or racist. One participant who was curious about the results of the study followed up with us. Upon hearing that the the black agents were slightly preferred over the white agents and not the other way around, the participant commented that if this study had been conducted in 2019, the results would be different. The participant attributed this preference of the black avatar to the recent incidents in the USA in the year of 2020. Another participant mentioned to us that "*we are in an extremely isolated and dangerous environment. I didn't even (want to) pay attention to the AI's social class and culture.*"

Third, the context for our studies is assisted decision making unlike prior work that focused on simulated virtual human agents in medical setting [42], and pictures of colored robots in shooter bias paradigm [4]. Moreover, in expert instructional role, black agents have shown to cause higher transfer of leaning and enhanced focus for students compared to white agents raising speculations that black agents may have warranted more attention because of being perceived as "novel" [5]. Thus, further research is required to understand the presence or absence of "implicit" social biases when the virtual avatars are portrayed as assistants in human-agent co-decision making. Fourth, our study population was skewed toward college students and young professionals; 49 participants were either in college or had college or higher degrees (the other 10 participants had high school diploma or equivalent certificates). The average age of the participants is 23 years. It has been suggested that younger populations tend to have lesser racial biases [14] than older generations. Moreover, our participants are mostly people of color (about only 19% Caucasian), who may have lower biases towards people of color [14].

5.3 Design Implications

5.3.1 Designing persuasive embodied virtual agents. Overall, our results are in line with the previous findings in HRI, indicating that the employment of strategic verbal and non-verbal cues can be effective in achieving persuasiveness [1, 15, 19, 49]. Embodied virtual agents that are designed to possess high practical knowledge and disseminate information fluently through well-structured and organized sentences from the user's perspective can be successful in obtaining user compliance while making a good impression. However, researchers should be mindful of the inherent limited fidelity of avatars since IVAs are lower on the embodiment spectrum than physical robots and humans. The strategies proven effective for robots and humans may not translate to digital agents completely. Further explorations are required to confirm the effectiveness of various vocal and non-verbal cues for creating persuasive IVAs.

5.3.2 Designing diverse agents for the digital world. Cave et al. pointed out that real and imagined humanoid robots, chatbots, and virtual assistants, as well as portrayals of AI in films and on the Internet are predominantly portrayed as white [13]. Search engines prioritize anthropomorphic images of AI that are "White" [33]. They further implied that imagining intelligent, professional, or powerful AI machines essentially means imagining ethnically "White" machines [13]. The construction of robots also mostly utilizes white materials and surfaces [48]. As such, researchers are raising concerns and questions on how AI mimics and reinforces knowledge that serves white supremacy [27]. Indeed, creating diverse embodied IVAs for the digital world is beneficial; a study on racial mirroring effect in psychotherapeutic conversations concluded that participants were eager to disclose more information to the agent from a different ethnic group even though they showed higher satisfaction, more closeness, and higher desire to continue interaction with same-ethnicity agents [31].

Contrary to previous findings and our hypotheses, we found no significance preference for dark or light skin-toned agent regardless of their expertise. This finding brings our attention to the ever-changing and complex social construct—ethnicity. Our results suggest that designers and developers should carefully consider the design of virtual agents as the appearance, voice, body language, and more may affect how these agents are perceived by users.

5.4 Limitations and Future Work

This work has several limitations that point to future research directions. First, skin tone is much more refined than just black and white. In this study, we only tested skin tone as black/dark or white/ light. However, skin tone and racial dynamics are far more diverse [38]. Within same ethnic and cultural groups, there are biases associated with different skin colors [25]. We also did not control for ethnicity of participants in this study which could have resulted in undetected in-group and out-group biases. Moreover, ethnicity is a social construct whereas skin tone is a biological trait; these two are often mixed without nuanced discussion. Future work should explicitly look into the influence of skin tone and other aspects of ethnicity on the perception of ethnicity of and implicit biases towards IVAs. Second, other aspects of agent appearance, such as hair styles, clothing, and voice, might have influenced the results of this study. For example, prior research indicated that other aspects of appearance such as attractiveness [28] or language [50] could influence the persuasiveness of agents. More research is needed to explore the complex design space of agent appearance to create diverse personas.

Third, future work should explore the interplay of expertise and implicit bias in different task contexts such as high-stakes, time pressured scenarios. For instance, prior work with medical students highlighted lesser empathy towards black virtual human agent [42], which calls for careful design of diverse agents to reduce reflection of such biases. Finally, our participants do not represent the full diversity in our society. Thus, the results presented in this paper should be understood and interpreted with this context. While we did not see significant effects of the participants' self-identified ethnicity on our results, future work should employ better ways to measure implicit biases towards agents who are perceived to be ethnically different.

Overall, future research should explore how various aspects of societal biases affect human-agent interactions and how the interplay between these variables may help create inclusive systems without reinforcing negative impacts of biases associated with the stereotypes.

ACKNOWLEDGMENTS

This work was partially supported by the National Science Foundation award #1840088 and the Johns Hopkins University Institute for Assured Autonomy. We thank Jaimie Patterson for proofreading this paper.

REFERENCES

[1] Sean Andrist, Erin Spannan, and Bilge Mutlu. 2013. Rhetorical robots: making robots more effective speakers using linguistic cues of expertise. In *2013 8th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. IEEE, 341–348.

[2] Domna Banakou, Parasuram D Hanumanthu, and Mel Slater. 2016. Virtual embodiment of white people in a black virtual body leads to a sustained reduction in their implicit racial bias. *Frontiers in human neuroscience* 10 (2016), 601.

[3] Christoph Bartneck, Dana Kulić, Elizabeth Croft, and Susana Zoghbi. 2009. Measurement instruments for the anthropomorphism, animacy, likeability, perceived intelligence, and perceived safety of robots. *International journal of social robotics* 1, 1 (2009), 71–81.

[4] Christoph Bartneck, Kumar Yogeeswaran, Qi Min Ser, Graeme Woodward, Robert Sparrow, Siheng Wang, and Friederike Eyssel. 2018. Robots and racism. In *Proceedings of the 2018 ACM/IEEE international conference on human-robot interaction*. 196–204.

[5] AL Baylor and Yanghee Kim. 2004. Pedagogical agent design: The impact of agent gender, ethnicity, and instructional role. (2004).

[6] Amy L Baylor. 2009. Promoting motivation with virtual agents and avatars: role of visual presence and appearance. *Philosophical Transactions of the Royal Society B: Biological Sciences* 364, 1535 (2009), 3559–3565.

[7] Timothy Bickmore and Justine Cassell. 2001. Relational agents: a model and implementation of building user trust. In *Proceedings of the SIGCHI conference on Human factors in computing systems*. 396–403.

[8] Sue Black and Eilidh Ferguson. 2011. *Forensic anthropology: 2000 to 2010*. CRC Press.

[9] Judee K Burgoon, Thomas Birk, and Michael Pfau. 1990. Nonverbal behaviors, persuasion, and credibility. *Human communication research* 17, 1 (1990), 140–169.

[10] Judee K Burgoon, Michelle L Johnson, and Pamela T Koch. 1998. The nature and measurement of interpersonal dominance. *Communications Monographs* 65, 4 (1998), 308–335.

[11] Estelle Campione and Jean Véronis. 2002. A large-scale multilingual study of silent pause duration. In *Speech prosody 2002, international conference*.

[12] Rolf Carlson, Kjell Gustafson, and Eva Strangert. 2006. Cues for hesitation in speech synthesis. In *Ninth International Conference on Spoken Language Processing*.

[13] Stephen Cave and Kanta Dihal. 2020. The whiteness of AI. *Philosophy & Technology* 33, 4 (2020), 685–703.

[14] Tessa ES Charlesworth and Mahzarin R Banaji. 2019. Patterns of implicit and explicit attitudes: I. Long-term change and stability from 2007 to 2016. *Psychological science* 30, 2 (2019), 174–192.

[15] Vijay Chidambaram, Yueh-Hsuan Chiang, and Bilge Mutlu. 2012. Designing persuasive robots: how robots might persuade people using vocal and nonverbal cues. In *Proceedings of the seventh annual ACM/IEEE international conference on Human-Robot Interaction*. 293–300.

[16] Patricia G Devine, Patrick S Forscher, Anthony J Austin, and William TL Cox. 2012. Long-term reduction in implicit race bias: A prejudice habit-breaking intervention. *Journal of experimental social psychology* 48, 6 (2012), 1267–1278.

[17] Friederike Eyssel and Frank Hegel. 2012. (s) he's got the look: Gender stereotyping of robots 1. *Journal of Applied Social Psychology* 42, 9 (2012), 2213–2230.

[18] Brian J Fogg. 1998. Persuasive computers: perspectives and research directions. In *Proceedings of the SIGCHI conference on Human factors in computing systems*. 225–232.

[19] Aimi S Ghazali, Jaap Ham, Emilia I Barakova, and Panos Markopoulos. 2018. Effects of robot facial characteristics and gender in persuasive human-robot interaction. *Frontiers in Robotics and AI* 5 (2018), 73.

[20] Michael Glass, Jung Hee Kim, Martha W Evens, Joel A Michael, and Allen A Rovick. 1999. Novice vs. expert tutors: A comparison of style. In *MAICS-99, Proceedings of the Tenth Midwest AI and Cognitive Science Conference*. 43–49.

[21] Rosanna E Guadagno, Jim Blascovich, Jeremy N Bailenson, and Cade McCall. 2007. Virtual humans and persuasion: The effects of agency and behavioral realism. *Media Psychology* 10, 1 (2007), 1–22.

[22] Belinda Gutierrez, Anna Kaatz, Sarah Chu, Dennis Ramirez, Clem Samson-Samuel, and Molly Carnes. 2014. "Fair Play": a videogame designed to address implicit race bias through active perspective taking. *Games for health journal* 3, 6 (2014), 371–378.

[23] Elin Johanna Hartelius. 2008. *The Rhetoric of Expertise*. Ph.D. Dissertation. The University of Texas at Austin.

[24] Mojgan Hashemian, Ana Paiva, Samuel Mascarenhas, Pedro A Santos, and Rui Prada. 2019. The power to persuade: a study of social power in human-robot interaction. In *2019 28th IEEE International Conference on Robot and Human Interactive Communication (RO-MAN)*. IEEE, 1–8.

[25] Margaret Hunter. 2007. The persistent problem of colorism: Skin tone, status, and inequality. *Sociology compass* 1, 1 (2007), 237–254.

[26] Marie Jarrell, Reza Ghaiumy Anaraky, Bart Knijnenburg, and Erin Ash. 2021. Using Intersectional Representation & Embodied Identification in Standard Video Game Play to Reduce Societal Biases. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. 1–18.

[27] Yarden Katz. 2020. *Artificial Whiteness: Politics and Ideology in Artificial Intelligence*. Columbia University Press.

[28] Rabia Fatima Khan and Alistair Sutcliffe. 2014. Attractive agents are more persuasive. *International Journal of Human-Computer Interaction* 30, 2 (2014), 142–150.

[29] J. C. Lafferty, Eady, and J. Elmers. 1974. *The desert survival problem*. Plymouth, Michigan: Experimental Learning Methods.

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

[31] Yuting Liao and Jiangen He. 2020. Racial mirroring effects on human-agent interaction in psychotherapeutic conversations. In *Proceedings of the 25th International Conference on Intelligent User Interfaces*. 430–442.

[32] Rosemarijn Looije, Mark A Neerincx, and Fokie Cnossen. 2010. Persuasive robotic assistant for health self-management of older adults: Design and evaluation of social behaviors. *International Journal of Human-Computer Studies* 68, 6 (2010), 386–397.

[33] Mykola Makhortykh, Aleksandra Urman, and Roberto Ulloa. 2021. Detecting race and gender bias in visual representation of AI on web search engines. In *International Workshop on Algorithmic Bias in Search and Recommendation*. Springer, 36–50.

[34] Jand Noel. 1999. On the varieties of phronesis. *Educational philosophy and theory* 31, 3 (1999), 273–289.

[35] Tatsuya Nomura. 2017. Robots and gender. *Gender and the Genome* 1, 1 (2017), 18–25.

[36] Brian A Nosek, Mahzarin R Banaji, and Anthony G Greenwald. 2002. Harvesting implicit group attitudes and beliefs from a demonstration web site. *Group Dynamics: Theory, Research, and Practice* 6, 1 (2002), 101.

[37] Kristine L Nowak and Frank Biocca. 2003. The effect of the agency and anthropomorphism on users' sense of telepresence, copresence, and social presence in virtual environments. *Presence: Teleoperators & Virtual Environments* 12, 5 (2003), 481–494.

[38] Ihudiya Finda Ogbonnaya-Ogburu, Angela DR Smith, Alexandra To, and Kentaro Toyama. 2020. Critical race theory for HCI. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. 1–16.

[39] Brigitte Planken. 2005. Managing rapport in lingua franca sales negotiations: A comparison of professional and aspiring negotiators. *English for Specific Purposes* 24, 4 (2005), 381–400.

[40] Irene Rae, Leila Takayama, and Bilge Mutlu. 2012. One of the gang: supporting in-group behavior for embodied mediated communication. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 3091–3100.

[41] Irene Rae, Leila Takayama, and Bilge Mutlu. 2013. The influence of height in robot-mediated communication. In *2013 8th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. IEEE, 1–8.

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

[43] Maike AJ Roubroeks, Jaap RC Ham, and Cees JH Midden. 2010. The dominant robot: Threatening robots cause psychological reactance, especially when they have incongruent goals. In *International Conference on Persuasive Technology*. Springer, 174–184.

[44] Ameneh Shamekhi, Q Vera Liao, Dakuo Wang, Rachel KE Bellamy, and Thomas Erickson. 2018. Face Value? Exploring the effects of embodiment for a group facilitation agent. In *Proceedings of the 2018 CHI conference on human factors in computing systems*. 1–13.

[45] Annette N Shelby and N Lamar Reinsch Jr. 1995. Positive emphasis and you-attitude: An empirical study. *The Journal of Business Communication (1973)* 32, 4 (1995), 303–326.

[46] Mikey Siegel, Cynthia Breazeal, and Michael I Norton. 2009. Persuasive robotics: The influence of robot gender on human behavior. In *2009 IEEE/RSJ International Conference on Intelligent Robots and Systems*. IEEE, 2563–2568.

[47] Stefanie Simon, Emily Shaffer, Rebecca Neel, and Jenessa Shapiro. 2019. Exploring Blacks' perceptions of Whites' racial prejudice as a function of intergroup behavior and motivational mindsets. *Social Psychological and Personality Science* 10, 5 (2019), 575–585.

[48] Robert Sparrow. 2019. Do robots have race?: Race, social construction, and HRI. *IEEE Robotics & Automation Magazine* 27, 3 (2019), 144–150.

[49] Tzu-Yang Wang, Ikkaku Kawaguchi, Hideaki Kuzuoka, and Mai Otsuki. 2018. Effect of Manipulated Amplitude and Frequency of Human Voice on Dominance and Persuasiveness in Audio Conferences. *Proceedings of the ACM on Human-Computer Interaction* 2, CSCW (2018), 1–18.

[50] Langxuan Yin, Timothy Bickmore, and Dharma E Cortés. 2010. The impact of linguistic and cultural congruity on persuasion by conversational agents. In *International Conference on Intelligent Virtual Agents*. Springer, 343–349.