Marcus or Mira - Investigating the Perception of Virtual Agent Gender in Virtual Reality Role Play-Training

Georg Regal
georg.regal@ait.ac.at
AIT Austrian Institute of Technology
Vienna, Austria

Jakob Uhl
jakob.uhl@ait.ac.at
AIT Austrian Institute of Technology
Vienna, Austria

Anna Gerhardus
gerhardus@ihs.ac.at
IHS Institute for advanced studies
Vienna, Austria

Stefan Suette
stefan.suette@ait.ac.at
AIT Austrian Institute of Technology
Vienna, Austria

Elisabeth Frankus
frankus@ihs.ac.at
IHS Institute for advanced studies
Vienna, Austria

Julia Schmid
schmid@ihs.ac.at
IHS Institute for advanced studies
Vienna, Austria

Simone Kriglstein
kriglstein@mail.muni.cz
Masaryk University
Brno, Czech Republic
AIT Austrian Institute of Technology GmbH
Vienna, Austria

Manfred Tscheligi
manfred.tscheligi@plus.ac.at
University of Salzburg
Salzburg, Austria
AIT Austrian Institute of Technology GmbH
Vienna, Austria

ABSTRACT

Immersive virtual training environments are used in various domains. In this work we focus on role-play training in virtual reality. In virtual role-play training conversations and interactions with virtual agents are often fundamental to the training. Therefore, the appearance and behavior of the agents plays an important role when designing role-play training. We focus on the gender appearance of agents, as gender is an important aspect for differentiation between characters. We conducted a study with 40 participants in which we investigated how agents gender appearance influences the perception of the agents´ personality traits and the self-perception of a participants' assumed role in a training for social skills. This work contributes towards understanding the design-space of virtual agent design, virtual agent gender identity, and the design and development of immersive virtual reality role-play training.

CCS CONCEPTS

• **Human-centered computing** → *HCI theory, concepts and models.*

KEYWORDS

Virtual Reality, Training, Gender, Virtual Agents

ACM Reference Format:
Georg Regal, Jakob Uhl, Anna Gerhardus, Stefan Suette, Elisabeth Frankus, Julia Schmid, Simone Kriglstein, and Manfred Tscheligi. 2022. Marcus or Mira - Investigating the Perception of Virtual Agent Gender in Virtual Reality Role Play-Training . In . ACM, New York, NY, USA, 11 pages. https://doi.org/10.1145/3562939.3565629

VRST2022, Nov29–Dec1, 2022, Tsukuba, Japan

ACM ISBN 978-1-4503-9889-3/22/11.
https://doi.org/10.1145/3562939.3565629

1 INTRODUCTION

Immersive virtual training environments have become increasingly important in recent years due to improvements in virtual reality (VR) hardware and software, the so-called "second wave of VR" [3]. Fully immersive virtual training environments offer certain benefits, as the trainees can experience highly realistic, lifelike scenarios that make them feel present in the virtual world and transfer learned skills into the real world. Also, immersive virtual training environments provide the possibility to manipulate and interact with the environment and allow to repeat scenarios - in the exact same configuration - as often as needed. Furthermore, virtual training environments allow to train in environments that are not available (e.g. environments that are not yet built [53]) or dangerous (e.g. blasting works in underground coal mines [27]). Thus, immersive virtual training applications are increasingly used in many fields, for example, medicine [8], emergency and safety training [27], maintenance [25, 42], police training [13, 44] and military training [1, 6]. Moreover, a broad range of existing training applications are aimed toward training soft skills or social skills, e.g. training against prejudice [17], mindfulness training [35], leadership development training [2], job interview training for people with mental illness [32], training on interacting with autistic people [38], perspective change to understand Paralympic sports [37], training embodied cultural interaction [16] and many more.

When designing a social skills training, the conversations and interactions with virtual agents are important, e.g. a trainee talking to a fictional interviewer in a job interview training, or a trainee negotiating with agents in a negotiation training. Also for social skills role-play training VR environments are promising as they offer high immersion while providing controlled repeatability.

Our decisions in social situations depend on how we understand the situation and our conversational partners. Similarly, the assumptions we make of a conversational agent in a role-play training might influence our decisions in such a training. In the appraisal of social situations, we argue that it is often attributions of personality traits and certain characteristics, that form our impression of others

and guide the decisions we make in conversations. These assumptions about personality are partly based on the appearance of our conversational partner, as Naumann et al. illustrated [43]. Therefore we investigate the associations of agent appearance and attributed personality traits and characteristics. However, appearance consists of many factors that can be considered when designing agents.

We focus on the gender appearance of the agent (cf. Figure 1 - - right side). It is important to notice that the visual appearance, voice, name and behaviour of the agent leads to the perception of the agent as a female* or male* body, following the theories on gender construction and performativity of gender [10, 62]. We like to emphasize that gender goes beyond biological sex and is based on social factors and social influences. We use the terms "female*" and "male*" and men*/women* to acknowledge the diversity in the gender spectrum. The * symbolizes that gender is not a binary construct but a spectrum which also includes people who identify as intersexual, non-binary and other non-tradition forms of gender performance or appearance. However, when we discuss related work we use similar terms as the authors (e.g. male/female and men/women).

In this work we investigate the influence of agent gender in an immersive role-play training for social skills [cf. 3.2]. To investigate the influence of agent gender in role-play training we have created two characters, that are perceived/coded as male* and female* (gender appearance) due to their visual appearance, their names and their voices. In a controlled user study, we investigate the following research questions and hypotheses:

RQ1 - Agent Personality Traits: Are different, gender-role-specific personality traits attributed to a male* and a female* coded virtual agent as counterpart? *Hypothesis 1:* Despite exactly the same scenario and conversation content, the attributed external personality and gender stereotypical characteristics of the virtual agents differ.

RQ2 - Self-perception of assumed Role: Does the self-perception of the adopted role differ in regard to the gender of the agent/the participant's own gender? *Hypothesis 2:* The attribution of personality traits and masculinity/femininity of the assumed role do not differ depending on the participants´ gender or the agents´ gender.

This work contributes towards understanding how agent gender influences perceived personality traits and gender stereotypes in immersive virtual training applications.

2 RELATED WORK

2.1 Doing and Undoing Gender

When examining the aforementioned research questions, we believe it is necessary to consider feminist perspectives on technology and its development. While early feminist researchers of technology pointed out the (under)representation of women in the development and usage of technology, and technology itself as being coded as masculine [29], later so-called 'Cyberfeminism' viewed the development of the internet and other technologies as a means to overcome the binary boundaries of genders (cf. Hawthorn and Klein [30], Sadie Plant [47] and Haraway [28]).

Gender can be understood as socially produced: With their concept of "doing gender", West and Zimmerman [62] referred to the underlying and repetitive process of production and performance of gender. Gender, following this approach, is understood as a product of social action that is constantly produced and reproduced in social situations and processes. Emphasizing the performative nature of gender, Butler [10] refers to the repeated rehearsal of existing gendered norms, conventions, and discourses and argues for the performativity of gender.

Moreover, socially learned roles are often continued in the digital environment. According to van Reijmersdal et al.[58], girls prefer avatars that are similar to them physically or character-wise and thus try to reconstruct their own identity in the game. Accordingly, women tend to play more sexy, beautiful, and strong characters and fewer characters that are perceived as aggressive. This can be understood as social conditioning, as it can be found in many other social practices. In a similar direction, Eklund [20] suggests that female players consciously decide the gender and design of their avatar in the online game World of Warcraft (WoW). In order to strengthen the identification with the character as much as possible, female players tend to choose a female avatar, as male avatars often do not fit their body image. Also, female players often use traditionally female connotative avatars, such as healing or supportive characters. In contrast, most men indicate selecting avatars that have the most utility for them. Similarly, Bergstrom et al. [5] could show that gender and profession of NPCs correlate in WoW and that gender stereotypes are reproduced in this virtual world.

Therefore, we aimed to analyze how such mutual shaping of gender and technology influences the perception of an agent and the agents´ personality traits in role-play training in immersive virtual reality. This has lead to the creation of the research question *RQ1 - Agent Personality Traits*.

Yee & Bailenson [63] pronounce a reflexive function of avatar identification, referring to the so-called "Proteus effect", which describes that the appearance of an avatar has an influence on the self-image and gameplay behavior of the player. Fox et al. [24] could show that a hypersexualized avatar has an influence on the gaming behavior and the self-perception of the person playing the game. Avatar gender has an impact on in-game behavior as noted by [4, 64]. This impact on in-game behavior - also by other players - is one of the reasons why gender swapping occurs in games.

We were interested in the self-perception of the participants especially which gender participants perceived for their virtual representation. Although participants did not see a visual representation of their avatar due to the VR first-person perspective, related work regarding avatar gender is relevant for research question *RQ2 - Self-perception of assumed role*. Thus, we administered the TMF scale [34] (see Section 3.3.2) as a control variable.

2.2 Virtual Agent and NPC Appearance

The design of virtual agents and non-playing characters (NPCs) is an influential aspect for the user experience when interacting with virtual characters.

Rogers et al. proposed [52] *"a methodological approach towards the visual design of NPCs to fit specific narrative roles"*. Chen [15] studied the influence of identification with NPCs (and player characters) and post-game attitude in a serious game about migration. Ogier [45] used methods from ethnography to analyze NPC appearance in the game Fallout 4. For NPCs as enemies in combat

Figure 1: Office Environment (left) and the two virtual agents: Marcus (middle) and Mira (right).

or shooter games, Rivera et al.[50] proposed design patterns and investigated the effects of NPC design on gameplay. Pimentel et al. [46] investigated affective/cognitive responses to customized NPCs and Carrigan et al.[12] found that the visual level of aggression (color, armor, and weapons) of an enemy crowd had a direct impact on players combat performance. Although in this work we focus on gender appearance it is important to consider the aforementioned related work especially when considering an intersectional approach [11] - as other aspects of agent design can be confounding variables that influence the overall experience apart from the avatar's gender.

Regarding NPC gender, Rogers et al. [51] found that NPC biological sex in combination with participants' gender had an influence on mentor and enemy identification, with higher values in the opposite biological sex condition. In a similar direction Feng et al. [22] systematically researched gender and proximity. In the described study, a virtual agent - male or female - provided negative feedback, while either standing still or approaching the participant. Results "*indicate that participants attribute greater self-blame (internal attribution) for their purported poor performance when interacting with the female virtual instructor than when interacting with the male virtual instructor*"[22, p810]. As a possible explanation Feng et al. suggests that the female instructor is perceived as *"less threatening and as a consequence less blameworthy"*[22, p816]. Moreover, a more positive affect in response was found for the female over the male instructor. Rapuano et. al. [48] investigated interpersonal distance between participants and agents in VR simulations, among other findings they could show that male participants preferred shorter distances to smiling female agents and female participants larger distances to angry male agents. In a similar direction Zibrek et al. [65] investigated proximity-based differences in motion pattern attractiveness and motion pattern gender cues. They found that attractiveness for motions influence proximity but no effect for gender could be found.

Silvervarg et al. [55] investigated the effects of visual gender appearance of embodied conversational agents, focusing on verbal abuse. Their results show that female conversational agents are more abused than male conversational agents and androgynous conversational agents are less abused than female conversational agents but more abused than male conversational agents. In a similar direction, Nag and Yalçın [41] investigated stereotypes towards virtual agents comparing female, androgynous, and male agents with few differences in their visual attributes. Their findings suggest a gender-balanced view of stereotypical traits. Female, androgynous and male agents were perceived as similar in competency, agency, and communion traits. However, the authors observed an interaction effect between participants' gender and the agents' gender in communion scores. Lucas et al. [36] compared personal trainer feedback given by humans, virtual agents, or avatars. Differences were found especially for negatively valenced messages, which were better regarded when given by a real human person than avatars and agents. Feijóo-García et al. [21] analyzed the effects of agents' gender (and accent) on mental wellness support. The authors could not find any difference in perceived expertise regarding the agents' gender and 50% of participants indicate no preference, however, a female-gendered agent was preferred by 37.93% while only 6.9% preferred a male agent - and 5,17% preferred neither.

Building on the work by [21, 22, 36, 41, 48, 51, 55, 65] we investigate if influences on the self-perception of the participant in relation to the agents´ gender can be found in a VR role play training, to be more precise a negotiation situation between the participant and a male* (Marcus) or female* (Mira) agent.

The related work on virtual agent appearance and gender and the interplay with user experience influenced the research question *RQ1 - Agents´ Personality Traits*, inspired by [5, 21, 36, 41, 45, 55] and the research question *RQ2 -Self-perception of assumed role* - inspired by [12, 15, 22, 51]

3 USER STUDY

3.1 Study Design

We conducted a between-subjects experiment with two independent variables: participants' gender (self identified as male* or female*) and agent gender (designed to be perceived as male* or female*). Agent gender was counter-balanced assigned to both male* and female* participants. We choose a between-subject design, as a within-subject design could make people aware of the study purpose and influence the outcome [7].

3.1.1 Participants. Forty participants were recruited via an online posting of the invitation to the experiment, as well as through an internal database. 21 self-identified as female*, 19 self-identified as male*. In our sample, ages ranged from 21 to 73, with a mean age of 41.3 (SD = 13.91). Education - International Standard Classification

of Education (ISCED) [23] - ranged from those who had only completed mandatory school (N = 2, ISCED 0-2), had completed some secondary education (N = 13, ISCED 3 - 4), to those who had some form of tertiary education (N = 25, ISCED 5 - 8).

3.1.2 COVID-19 Precautions. This study was conducted during the COVID-19 pandemic, thus precautions were taken. All participants and facilitators wore FFP2 masks during the study. Participants could take off the masks when wearing the VR headset to prevent fogging of the lenses. All items (VR headset, pens, table, etc.) were disinfected before and after the study.

3.1.3 Apparatus. The study was done in the laboratory of AIT Austrian Institute of Technology. The study was conducted using an HTC Vive Pro Eye VR headset [59] and an Alienware Area 51 Gaming PC [19] with the following specifications: Intel i7-58290K 3.3GHZ, NVIDIA Geforce 1080TI, 16GB RAM, . The role-play training was developed in Unity [57] and used the Microsoft Azure Speech API [40] for implementing the speech interface.

3.1.4 Procedure. The study started with information about COVID-19 precaution measures (e.g. the disinfection strategy). All participants signed an informed consent.

Afterwards participants received an introduction to the VR headset and the controllers, equipped the VR headset, and experienced a peaceful mountain house (the standard Steam VR [18] welcome room) to get familiar with being in a VR environment. When participants were ready the training scenario was started (see Section 3.2).

After the VR experience, participants were asked to fill out the following questionnaires (see Section 3.3), provided on a tablet and implemented in the software LimeSurvey [26] : First four questionnaires about their impression of the agent: (1) BFI-2-S [56] for the agent (2) TMF [34] for the agent, (3) Uncanny Valley-Questionnaire (based on [49]) for the agent and (4) Agency/Communality semantic differential (based on [31]) for the agent. Secondly, two questionnaires for their assumed executive role: (1) BFI-2-S [56] for their own role and (2) the TMF scale [34] for their own role. Afterwards a short interview was conducted. Participants were asked which 5 adjectives they would assign for the agent (Mira or Marcus). The study ended with a debriefing and receiving the remuneration.

3.2 Scenario

The scenario was produced to be as realistic as possible. The story of the scenario was developed based on user research with middle management employees within a large organization. The main theme that guided the creation of the scenario was the need to "learn to say no in a respectful and professional way". The scenario was developed in strong collaboration between a psychologist who conducted the user research and a dramaturgist for developing a coherent story. Also in terms of technical development, we aimed for high realism through motion capturing of the agents' movements and voice enactment by hired actors. The same motion capturing data was used for Mira and Marcus, performed by a female* actor, as motion patterns can influence the perception of an agent [65].

All participants experienced the same scenario, described below. Only the gender appearance -visual appearance, voice, and name - of the agent (Mira or Marcus) differed for the two groups.

The VR training experience is a role play in which the participant should be trained for social skills in a work context. Participants take on the role of a character from middle management in a large company. In the role-play training the participants interact in a dialogue with an agent.

The agents´ behavior and the conversation were completely the same only the gender appearance - visual appearance, the name, and the voice recordings - of the agent changed. Also, the conversation with the agent did not have any influence on the reaction of the agent nor the outcome of the scenario, which were the same for both agents, to keep the scenario similar and comparable for all participants.

The scenario started with an introduction to the story and the role of the participant. The participant was introduced to his/her role in the training, a middle management executive who is responsible for a team. The participant was told that they had to step in for a colleague in an important executive meeting that starts in ten minutes. At the same time, an appointment with a team member (agent - Mira or Marcus) is scheduled. As the spontaneous meeting is more important, the participant needs to cancel the meeting with the agent (Mira or Marcus). This introduction was displayed in text and in spoken by an invisible narrator.

The story introduction to the scene was followed by an introduction on how to use the controller and the speech interface. In the conversation with the agent, participants had to choose replies from a selection of possible answers, by speaking out loud one of the presented options.

After the introduction, the role-play started. The virtual agent (Mira or Marcus) approached the office (cf. Figure 1 - - left side), knocked on the door, and entered the room stating that he/she is ready for the meeting. In the enfolding conversation, the participant had to tell Mira/Marcus that there was no time for the meeting today. The whole situation then escalates as the meeting had already been postponed before, and Mira/Marcus feels being treated unfairly. In the end Mira/Marcus storms off and passive-aggressively states that she/he will not join the company trip next week.

3.3 Measures

We defined the following measurements: For *RQ1 - Agent Personality Traits* - we administered measurements to investigate the personality of the agent (see Section 3.3.1) and gender stereotypes assigned with the agent (see Section 3.3.1). Moreover, participants assigned five attributes to describe the agent. Similar to *RQ2 - Self-perception of assumed Role* - we administered measurements to investigate the perceived personality of the own role (see Section 3.3.1 and perceived gender stereotypes of the own role (see Section 3.3.1). As a control variable, we also included measurements for the uncanny valley effect (see Section 3.3.3).

3.3.1 Personality. The short form of the Big Five Inventory-2 (BFI-2-S) [56] was administered two times during the study: a) to assess the impression the agent made on the participants and b) to assess the perceived personality traits of the assumed executive role. The BFI-2-S consists of 30 items, which each count towards one of the five main personality factors (extraversion, agreeableness, conscientiousness, negative emotionality, and open-mindedness) and one of fifteen facet scales. The five-factor model of personality is one of

the most researched and well-accepted frameworks of personality [39], therefore we deemed it a fitting measurement instrument for our research question. The short version was chosen partly because of a faster administration time (30 items vs. 60 items), but also because participants were rating their own role and the agent based on relatively short interaction times of approximately ten minutes. For a valid description of the first impression, the level of detail provided by the short version was deemed sufficient. The BFI-S provides a good internal consistency of alpha ranging from α = 0.74 to α = 0.84 for the five main scales.

3.3.2 Gender Stereotypes. Two measurements were used to assess gender stereotypes towards the agent.

Traditional masculinity-femininity. Based on[10, 62] we explicitly did not tell people that Mira or Marcus should be seen as a female* / male* agent but used gender-neutral language during the study - e.g. referring to the agents by name or using terms like "the person you were negotiating with". To control whether the agents´ gender was attributed as intended (i.e. Mira as female* and Marcus as male*), we used the traditional masculinity-femininity survey (TMF) [34], which consists of six items that assess central facets of ascribed masculinity-femininity. The TMF was also used to assess the assumed masculinity or femininity of the assumed role. We removed three items from the TMF ("Ideally, I would like to be...", "Traditionally, my interests would be considered as...", "Traditionally, my attitudes and beliefs would be considered as..."), as they were not applicable for rating an agent or an assumed role instead of oneself. The TMF was used to rate the agent, as well as the own role. The TMF uses a 7-point Likert Scale, with 1 and 7 being the masculine and feminine poles respectively. With an internal consistency of α = 0.9, reliable measurements can be expected. The used items were: "I consider my avatar/ the agent to be ...", "Traditionally, the behavior of my avatar/ the agent would be considered ..." and "Traditionally, the physical appearance of my avatar/the agent would be considered ...".

Agency-Communality. An adapted form of the adjective lists of Hentschel et al. [31] was administered in the form of a semantic differential. The adjective lists were supplemented with their opposite to form the scales. Agency and communality are two dimensions that are sensitive to gender stereotypes in the form that agency (i.e. ability to act) is often associated with men*, whereas communality (sense of community) is more associated with women*. Agency consists of the four facets instrumental competence, leadership competence, assertiveness, and independence, communality consists of the three facets concern for others, sociability, and emotional sensitivity. Each of these sub-scales is rated with three to four items on a 7-point scale. The internal consistency of the seven main dimensions ranges from α = 0.75 to α = 0.91, indicating an adequate to good reliability.

3.3.3 Uncanny Valley Effect. As a control variable we documented the Uncanny Valley Effect of the agent, based on Reuten, van Dam & Naber [49], we used 4 (adapted) items for an assessment of uncanniness on a 5-point Likert Scale - perceived humaneness, naturalness, eeriness, and attractiveness of the agent.

4 RESULTS

4.1 Personality

We assessed the impression of personality traits for the agent and the perceived personality traits of the assumed executive role. The ratings of the big five personality traits of the assumed leadership role and of the agent were subjected to a two-way analysis of variance having two levels of participants' gender (male*/female*) and two levels of group (male* agent/female* agent).

4.1.1 Agent Personality traits. For the agent, no effects were statistically significant at the .05 significance level.

Upon investigating the sub-facets of conscientiousness, a significant effect of group was yielded for the factors organization (F(1, 36) = 6.11, p = 0.018) and productiveness (F(1, 36) = 4.64, p = 0.038). In both cases, participants rated the female* agent to be more organized (M = 4.35, SD = 0.56; M = 3.78, SD = 0.88) and more productive (M = 4.33, SD = 0.57; M = 3.83, SD = 0.86) than the male* agent. Subsequent t-tests resulted in a Cohen´s d of d = 0.78 for the facet organization and d = 0.68 for the factor productiveness, which indicates a medium to large effect size.

A significant effect of gender, as well as group, was yielded for the sub-facet assertiveness (main factor extraversion). Participants rated the female* agent to be more assertive (M = 3.53, SD = 0.73) than the male* agent (M = 3.03, SD = 0.87), with a Cohen´s d of d = 0.88, which constitutes a large effect. Also, female* participants rated their agent counterpart as more assertive (M = 3.6, SD = 0.80) than male* participants did (M = 2.92, SD = 0.73), independent of the agent´s gender, with a Cohen´s d of d = 0.62, a medium effect.

4.1.2 Assumed executive role personality traits. Gender yielded a significant effect for the factor 'negative emotionality' (F(1, 36) = 5.10, p = 0.030), with male* participants rating their assumed roles negative emotionality significantly lower (M=2.30, SD = 0.66) than female* participants (M=2.75, SD = 0.62). Upon further investigating the sub-facets of negative emotionality, it was found that the effect emerges from the scales 'anxiety' (F(1, 36) = 8.33, p = 0.007) with a Cohen's d of d = 0.82 and 'depression' (F(1, 36) = 6.59, p = 0.015), Cohen's d = 0.92, with male* participants rating their assumed role to be less anxious and depressed than female* participants. Looking at the sub-facets, a significant effect was yielded for 'trust' (F(1, 36) = 4.47, p = 0.041), a sub-facet of agreeableness, where participants who interacted with a male* agent rated their assumed role to be more trusting than when interacting with a female* agent. Cohen´s d for this group difference was d = 0.65, a medium effect size. Upon investigating the sub-facets of openness, a significant main effect of gender on intellectual curiosity was yielded (F(1, 36) = 4.57, p = 0.039), with male* participants rating their assumed role to be more open to ideas (M = 3.7, SD = 0.82) than female* participants (M= 3.2, SD = 0.86), with a Cohen´s d of d = 0.65.

4.2 Gender Stereotpyes

4.2.1 Agency and Communality. A two-way ANOVA was performed to analyze the effect of gender and group on the facets of the two constructs agency and communality. This revealed that there was no statistically significant main or interaction effect for communality. Main effects analysis showed, that gender of the participant did have a statistically significant effect on agency(F(1,38)

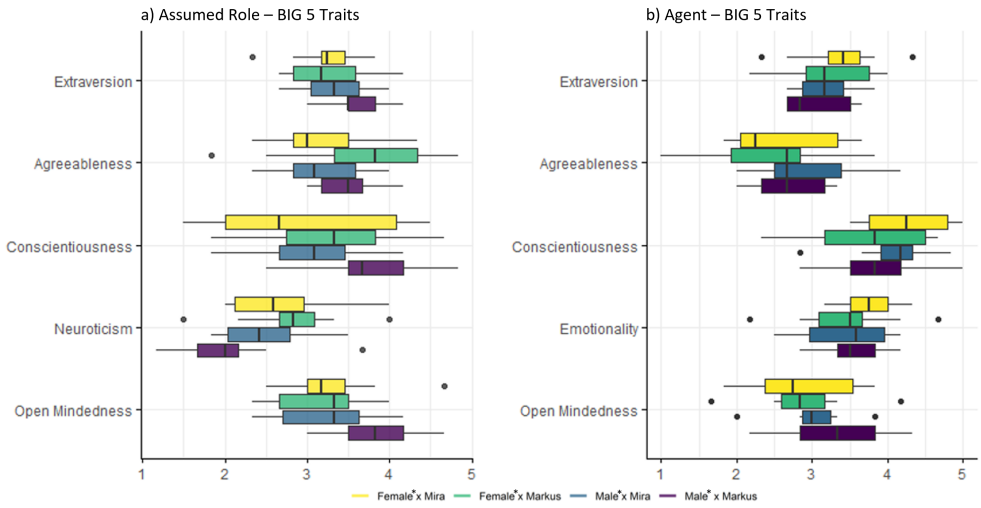

Figure 2: Boxplot BFI Factors of the own role (left), Boxplot BFI Factors of the agent (right). X-Axis denotes the mean of the scale, with items rated on a 5-point Likert scale.

Table 1: BFI: ANOVAs of selected sub facets for the virtual agent.

Parameter	Factor	SS	df	MS	F	p	Tukey	Cohens d
Assertiveness [Extraversion]	Group	2,86	1	2,86	5,31	0,027*	Mira >Marcus; Female* >Male*	d(group) = 0.88 d(sex) = 0.62
	Gender	**4,53**	**1**	**4,53**	**8,43**	**0,006***		
	Group x Gender	0,23	1	0,23	0,43	0,528		
	Error	19,36	36	0,54				
Organization [Conscientiousness]	**Group**	**3,38**	**1**	**3,38**	**6,11**	**0,018***	Mira >Marcus	d(group) = 0.78
	Gender	0,14	1	0,14	0,26	0,616		
	Group x Gender	0,64	1	0,64	1,15	0,291		
	Error	19,93	36	0,55				
Productiveness [Conscientiousness]	**Group**	**2,58**	**1**	**2,58**	**4,64**	**0,038***	Mira >Marcus	d(group) = 0.68
	Gender	0,20	1	0,20	0,37	0,549		
	Group x Gender	0,00	1	0,00	0,01	0,938		
	Error	19,99	36	0,56				

* indicates significant difference at $\alpha = 0.05$

= 6.87, p = 0.013). This effect resulted in a Cohen´s d of d = 0.82, a large effect between the two gender groups. Female* participants attributed more agency toward the VA(M = 4.83, SD = 0.16) than male* participants did (M = 4.25, SD = 0.15), regardless of the agents´ gender appearance.

4.2.2 Traditional masculinity-femininity scale (TMF). For the traditional masculinity-femininity scale and its subfacets, a two-way ANOVA was performed to analyze the effect of gender and group.

TMF Ratings Agent. For the ratings of the agent, a significant main effect of group was yielded for the TMF scale (F(1, 38) = 23.15, p <0.001) as well as for the subfacets. Participants rated the female* agent to be more feminine (M = 4.48, SD = 0.26) and the male* agent to be more masculine (M = 2.75, SD = 0.27), with a large effect size of d = 1.47.

TMF Ratings Assumed Executive Role. A significant main effect of participants gender could be detected for the TMF overall scale(F(1, 36) = 11.75, p = 0.002) when participants were rating their assumed executive role. Male* participants rated their role to be more masculine (M = 2.82, SD = 0.20), female* participants perceived their assumed role more neutral (M = 4.00, SD = 0.26), with a effect size of d = 1.12.

4.2.3 Uncanny Valley. As a control variable we documented the Uncanny valley effect for both agents. For the ratings of naturalness, humaneness, attractiveness and eerieness, as well as a combined score of uncannyness, a two-way analysis of variance having two levels of gender (male*/female*) and two levels of group (male* agent, female* agent) was performed. No significant main or interaction effects were found for the combined score of uncannyness, or the subfacets naturalness, humaneness or eerieness. A significant main effect of group was yielded for the factor attractiveness (F(1, 34) = 4.63, p = 0.038). A post hoc Tukey test showed that the attractiveness ratings of Marcus were significantly higher (M =2.6, SD = 1.1) than Mira's (M = 1.9, SD = 0.67), Cohen´s d of d = 0.67.

Table 2: BFI: ANOVAs of selected sub facets for the assumed role.

Parameter	Factor	SS	df	MS	F	p	Tukey	Cohens d
Intellectual curiosity [Openness]	Group	1,43	1	1,43	2,21	0,146	Male* >Female*	d(gender) = 0.65
	Gender	**2,95**	**1**	**2,95**	**4,57**	**0,039***		
	Group x Gender	2,15	1	2,15	3,33	0,076		
	Error	23,25	36	0,65				
Depression [Neuroticism]	Group	0,25	1	0,25	0,36	0,551	Female* >Male*	d(sex) = 0.82
	Gender	**4,62**	**1**	**4,62**	**6,59**	**0,015***		
	Group x Gender	0,57	1	0,57	0,82	0,373		
	Error	25,25	36	0,7				
Anxiety [Neuroticism]	Group	0,09	1	0,09	0,21	0,653	Female* >Male*	d(sex) = 0.92
	Gender	**3,65,**	**1**	**3,65**	**8,33**	**0,007***		
	Group x Gender	0,40	1	0,4	0,92	0,344		
	Error	15,80	36	0,44				
Trust [Agreeableness]	**Group**	**1,58**	**1**	**1,58**	**4,47**	**0,041***	Marcus >Mira, Female* & Marcus > Female* & Mira	d(group) = 0.65
	Gender	0,04	1	0,04	0,11	0,741		
	Group x Gender	**1,55**	**1**	**1,55**	**4,39**	**0,043***		
	Error	12,71	36	0,35				

* indicates significant difference at $\alpha = 0.05$

Table 3: Agency and Communality: ANOVAs of the two sub facets for the virtual agent.

Parameter	Factor	SS	df	MS	F	p	Tukey	Cohens d
Agency	Group	0,56	1	0,56	1,12	0,299	Female* >Male*	d(gender) = 0.82
	Gender	**3,46**	**1**	**4,46**	**6,87**	**0,013***		
	Group x Gender	0,12	1	0,02	0,04	0,846		
	Error	18,14	36	0,50				
Communality	Group	0,00	1	0,00	0,00	1,000		
	Gender	0,02	1	0,02	0,03	0,860		
	Group x Gender	0,01	1	0,01	0,01	0,910		
	Error	29,05	36	0,81				

* indicates significant difference at $\alpha = 0.05$

Table 4: Traditional Masculinity-Femininity: ANOVAS for the assumed role and the virtual agent.

Parameter	Factor	SS	df	MS	F	p	Tukey	Cohens d
Assumed role	Group	0,34	1	0,34	0,29	0,593	Male* <Female*	d(sex) = 1.12
	Gender	**13,60**	**1**	**13,60**	**11,75**	**0,002***		
	Group x Gender	0,04	1	0,04	0,04	0,851		
	Error	41,66	36	1,16				
Agent	**Group**	**30,04**	**1**	**30,04**	**23,15**	**<0,001***	Marcus <Mira	d(group) = 1.47
	Gender	4,58	1	4,58	3,53	0,069		
	Group x Gender	1,44	1	1,44	1,11	0,300		
	Error	46,73	36	1,30				

* indicates significant difference at $\alpha = 0.05$

5 DISCUSSION

5.1 RQ1 - Agent Personality Traits

Participants rated the gender of the agent on the TMF scale [34]. Results show that participants provided significantly different ratings regarding gender between Mira and Marcus, Mira was rated more feminine (M = 4,48, SD = 0,26) and Marcus was rated more masculine (M = 2,75, SD = 0,27). Noteworthy, on a 7-point scale, the rating (M = 4,48, SD = 0,26) for Mira is only slightly feminine, different from the ratings described by [34, Figure 1]. In general male* participants rated Mira´s gender as more coherent and Marcus´s gender with more deviation. Female*participants rated Marcus´s gender as more coherent and Mira´s gender with more deviation (Figure 4). This is important to consider as the subsequent interpretations of assumed personality and gender stereotypical characteristics might rely on this interpretation of Mira's gender.

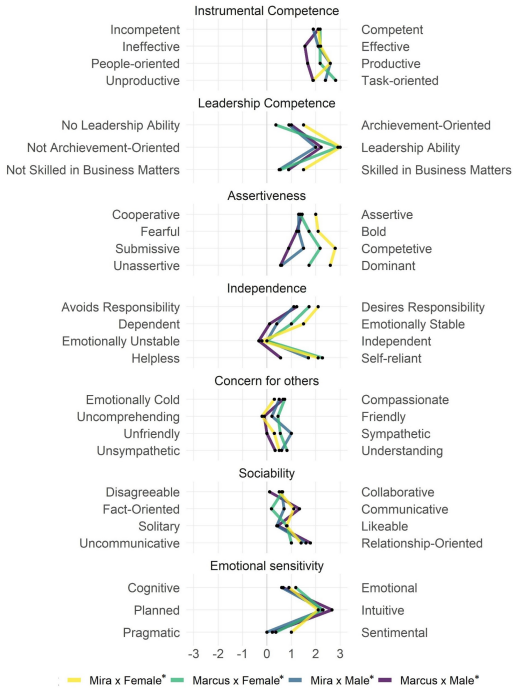

Figure 3: Agency and Communality profiles. X-Axis denotes means of the respective items on a 7-point scale ranging from -3 to +3.

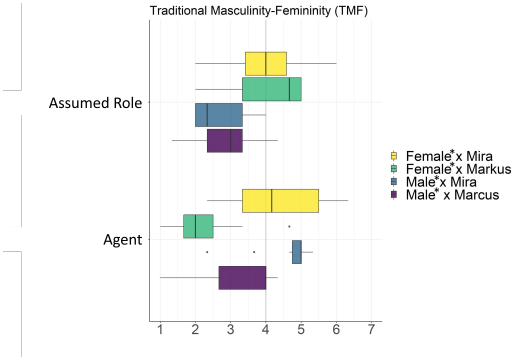

Figure 4: Boxplot of TMF Ratings. X-Axis denotes mean value for the respective group, with items on a 7-point Likert scale.

In the context of gendered stereotypes at the workplace organisational and management studies have made efforts to explain gendered behaviours. Burgess and Borgida [9] identify prescriptively (roles and characteristics that are assumed to be feminine/masculine) and descriptive (behaviour that is observed) gender stereotypes in organizations that influence both the way women and men are viewed in their position and how they perform their role. We assume that these are embedded in the unconscious knowledge of participants and are continued in the virtual environment [10, 61, 62]. Participants rating of Mira's fierce behaviour in the work environment can be contextualised in this, as we assume that the non-compliance with the gender stereotypes might have led to the "neutral" rating on the TMF scale.

Also, there is a difference in the spread in the TMF rating between male* and female* participants - cf. Figure 4 - with male* participants providing a more coherent rating for Mira's gender than female* participants. A possible explanation is that female* participants, who themselves experienced negative consequences due to such behaviour or who would not dare to act in such a way in a similar situation, considered this behaviour as less female* typical and thus more masculine*. This is also reflected in related work, in a meta-study on gendered stereotypes in management roles, Castanio et al. [14, p. 14] found that if women adopt male-coded behaviour women are perceived as "cold and instrumental" with does not apply for men the other way around.

Regarding **Hypothesis 1 - Despite exactly the same conversation content, the attributed external personality and gender stereotypical characteristics of the agents differ** - we could not find any significant differences for the BFI-2-S [56] main scales - neuroticism, extraversion, agreeableness, openness and conscientiousness. However, in the sub-facet assertiveness of the BFI-2-S a significant difference could be found, with Mira being rated as more assertive than Marcus. Also female* participants rate Mira different from male* participants. Participants rated Mira as significantly more organized and more productive than Marcus, in the BFI-2-S sub-facets organization and productiveness. In contrast, [7] found no effect for agent gender (and gender swap) on perceived gender traits of an agent, respective the played role.

Similarly, in the agency and communality semantic differential (adapted based on [31]) we could see a significant difference in agency. On the sub-scales, assertiveness and independence female* participants rated both agents as more assertive than men. Moreover, we could see an interaction effect that Mira is rated as more assertive than Marcus by female* participants. Also, female* participants interpreted the behavior of Mira as rather bold (see Figure 3). A possible explanation could be, that participants rated the behavior based on incorporated gender stereotypes that also apply to same-gender roles. Sheppard and Aquino [54] have shown that conflict between women is rated by observers as more severe than between men. This could explain why Mira's handling of the situation is rated differently than Marcus', although the content, body movement, and setup are the same.

Based on this, we assume that these gender stereotypes from real-life office environments are continued also in virtual environments, which is in line with studies that found such stereotypical behavior reenactment in games [4, 5, 20, 64] and in contrast to the utopia of a detachment of binary gender structures (man/woman) that cyberfeminism in the 90s postulated (cf. [60]).

Thus, we can not reject Hypothesis 1 as we could show that the stereotypical characteristics of the agent differ between a female* and a male* agent, even when acting completely similarly.

Therefore, when designing agents and especially when using virtual agents in role-play training it is important to consider the effects of agents' gender on the perception and the reaction of the user. As agents' appearance and gender identity have an influence [22, 48, 51, 55] on the users' behavior and preference [21]. This is especially important in role-play training e.g. negotiation situation [33] or when using agents to feedback [22] or consultation [21].

5.2 RQ2 - Self-perception of Role

As a control variable, we administered the TMF [34] scale to document how participants interpret the gender of their role. Male* participants rated their role as significantly more masculine than female* participants (M = 2,82, SD = 0,20) and female* participants rated their role as significantly more feminine than male* participants, but overall neutral (M = 4,00, SD = 0,26). A possible explanation could be that female* participants´ interpreted the behaviour of their role as more masculine, which is further supported by the spread in the TMF rating (cf. Figure 4). Existing gender stereotypes that prevail in the physical world, especially in the work context, are transferred to the assumptions of the own role during the VR experiment (see Wajcman [61]). Neither body nor voice of the experienced executive role was presented to the participants, yet only behaviour - given selected answers - during the conversation was the only cue that could be interpreted towards gender. Another possible interpretation could be that female* participants´ could identify less with the behaviour of their role. In the presented scene, participants incorporated the role of a manager - which is coded in Western society mainly as a male domain with matching gender stereotypes [14]. Women's traditional role in organisations is often seen as care-takers and in administrative tasks in organisations and the lack of leadership skills that is assigned to women. Although participants were not assigned a gender of the manager, female* participants did assume it neutral (which can be understood as either some see it as male* others as female* or neutral - cf. Figure 4). Especially for female* participants who, due to the behaviour of their role, considered their role as (more) masculine, similar effects (e.g. less embodiment, less agency and less ownership for female* participants in a male* avatar gender swap condition) as described by Bolt et al. [7] could have influenced participants self-perception and participants behaviour [63]. Male* participants however perceived their role as definitely male* (a bit stronger when interacting with Marcus than with Mira). Also on the BFI inventory male* participants rated their role significantly lower on the factor 'negative emotionality'. This might be an indication that male* participants see the behaviour of their role as more socially acceptable.

Regarding **Hypothesis 2 - The self-perception of the role in training does not differ depending on the gender of the study participant** there was a tendency, yet not significant on a 0.05 significance level, that male* participants - at least in this scenario - are more agreeable and open (BFI scales agreeableness and openness) in a fierce discussion when talking to a male* agent. A possible explanation could be that when talking to Mira the male* participants rather felt as if they (respectively their role) had made a mistake. In contrast, when talking to Marcus male* participants rather thought that Marcus was acting unnecessary emotional. This is supported by observations of the study facilitators as some participants mentioned these assumptions. The neutral TMF rating for Marcus by male* participants indicates this direction. Further support for this interpretation comes from the BFI-2-S sub-facets, as Mira was rated as more organized and more productive than Marcus, also the assigned attributes were more positive for Mira as well. This could have resulted in higher self-blame/internal attribution processes, which is in line with related work [22] who found higher self-blame for students that interact with a female professor compared to those interacting with a male professor.

Thus, we can reject Hypothesis 2 - as the self-perception of the assumed manager role depends on the gender of the participants and this might affect the role participants are performing in role-play training. Thus for the future design of role-play training (in a productive non-experimental setup) the role of participants should be defined and presented to the trainees as clearly as possible. This could be done by verbally telling participants details of their role (e.g. age, gender, if they perform themselves or somebody else, etc.). Alternatively, e.g. they could see their character in a mirror in the first-person perspective, or in non-VR setups their role can be seen from a third perspective.

5.3 Limitations and Future Work

The appearance and design of agents have many facets. In this study, we focus on the gender aspects in the design of two agents and the self-perception of gender during the VR experience. Still, there are many more aspects in the visual design that could have a further influence, that shall be investigated in the future. For example a hypersexualized agent (cf. [24]) and or different clothing of the agent could have resulted in different ratings. In this study, we intentionally decide to dress both agent´s (Mira and Marcus) quite similar (cf. Fig. 1), yet a stronger differentiation and maybe a more stereotypical female* clothing could have resulted in different ratings in the TMF scales and other scales. Marcus was rated as more attractive than Mira, thus this difference in attractiveness might be a further influential factor additional to the agents' gender. Also other appearance parameters like skin tone, hair color, name, etc. are likely to have an influence. This is especially important when considering a - widely agreed on (cf. [11]- intersectional approach, as not only gender by also other variables like ethnicity, social status, etc. might contribute towards the perception of agents. Thus in future work, we will investigate a broader range of agent designs, following an intersectional approach.

6 CONCLUSION

We conducted a study with 40 participants in which we investigated how different gender representations of agents (Mira and Marcus) in VR role play training affect the participants' experience. With our multi-dimensional perspective, which is not based only on VR literature but also on game, gender, and organizational studies' literature to contextualize our findings we aim to contribute to a broader understanding VR experiences.

This works contributes towards understanding how the agents´ gender representation affects (RQ1) the perception of agents´ personality traits, (RQ2) the self-perception of the participants' role. Nevertheless, future studies are needed to further investigate how agents´ appearance, gender, ethnicity, age, clothing, etc. influence agent perception, self-perception, and overall user experience.

ACKNOWLEDGMENTS

This work was partially funded by the Austrian Research Promotion Agency FFG under grant 872573. We would like to thank the project partners (institute for cultural excellence, Polycular and Karl Landsteiner Privatuniversität) for their support!

REFERENCES

[1] Kunjal Ahir, Kajal Govani, Rutvik Gajera, and Manan Shah. 2020. Application on virtual reality for enhanced education learning, military training and sports. *Augmented Human Research* 5, 1 (2020), 1–9.

[2] Mariano Alcañiz, Elena Parra, and Irene Alice Chicchi Giglioli. 2018. Virtual reality as an emerging methodology for leadership assessment and training. *Frontiers in Psychology* 9, SEP (2018), 1–7. https://doi.org/10.3389/fpsyg.2018.01658

[3] Christoph Anthes, Markus Wiedemann, and Dieter Kranzlmüller. 2016. State of the Art of Virtual Reality Technology. In *2016 IEEE Aerospace Conference*. IEEE, Big Sky, Montana, United States, 1–19. https://doi.org/10.1109/AERO.2016.7500674

[4] Kelly Bergstrom, Jennifer Jenson, and Suzanne de Castell. 2012. What's 'choice' Got to Do with It? Avatar Selection Differences between Novice and Expert Players of World of Warcraft and Rift. In *Proceedings of the International Conference on the Foundations of Digital Games* (Raleigh, North Carolina) *(FDG '12)*. Association for Computing Machinery, New York, NY, USA, 97–104. https://doi.org/10.1145/2282338.2282360

[5] Kelly Bergstrom, Victoria McArthur, Jennifer Jenson, and Tamara Peyton. 2011. All in a Day's Work: A Study of World of Warcraft NPCs Comparing Gender to Professions. In *Proceedings of the 2011 ACM SIGGRAPH Symposium on Video Games* (Vancouver, British Columbia, Canada) *(Sandbox '11)*. Association for Computing Machinery, New York, NY, USA, 31–35. https://doi.org/10.1145/2018556.2018563

[6] Olaf Binsch, Charelle Bottenheft, Annemarie Landman, Linsey Roijendijk, and Eric H.G.J.M. Vermetten. 2021. Testing the applicability of a virtual reality simulation platform for stress training of first responders. *Military Psychology* 33, 3 (2021), 182–196. https://doi.org/10.1080/08995605.2021.1897494

[7] Elena Bolt, Jasmine T Ho, Marte Roel Lesur, Alexander Soutschek, Philippe N Tobler, and Bigna Lenggenhager. 2021. Effects of a virtual gender swap on social and temporal decision-making. *Scientific Reports* 11, 1 (2021), 1–15.

[8] Kristina Bucher, Tim Blome, Stefan Rudolph, and Sebastian von Mammen. 2019. VReanimate II: training first aid and reanimation in virtual reality. *Journal of Computers in Education* 6, 1 (2019), 53–78. https://doi.org/10.1007/s40692-018-0121-1

[9] Diana Burgess and Eugene Borgida. 1999. Who women are, who women should be: Descriptive and prescriptive gender stereotyping in sex discrimination. *Psychology, public policy, and law* 5, 3 (1999), 665–692.

[10] Judith Butler. 2012. *Das unbehagen der Geschlechter*.

[11] Maria Carbin and Sara Edenheim. 2013. The intersectional turn in feminist theory: A dream of a common language? *European Journal of Women's Studies* 20, 3 (2013), 233–248.

[12] Emma Carrigan, Elena Kokkinara, F Gheorghe, M Houlier, Stéphane Donikian, and Rachel McDonnell. 2016. Crowd Appearance Affects Player Performance in Game Combat Scenarios. In *Proceedings of the 9th International Conference on Motion in Games* (Burlingame, California) *(MIG '16)*. Association for Computing Machinery, New York, NY, USA, 187–192. https://doi.org/10.1145/2994258.2994273

[13] Polona Caserman, Miriam Cornel, Michelle Dieter, and Stefan Göbel. 2018. A concept of a training environment for police using VR game technology. In *Joint International Conference on Serious Games*. Springer, 175–181.

[14] Ana M. Castaño, Yolanda Fontanil, and Antonio L. García-Izquierdo. 2019. "Why Can't I Become a Manager?"—A Systematic Review of Gender Stereotypes and Organizational Discrimination. *International Journal of Environmental Research and Public Health* 16, 10 (2019). https://doi.org/10.3390/ijerph16101813

[15] Vivian Hsueh Hua Chen and Wei Jie Dominic Koek. 2020. *Understanding Flow, Identification with Game Characters and Players' Attitudes*. Association for Computing Machinery, New York, NY, USA. https://doi.org/10.1145/3402942.3409784

[16] Alan Cheng, Lei Yang, and Erik Andersen. 2017. *Teaching Language and Culture with a Virtual Reality Game*. Association for Computing Machinery, New York, NY, USA, 541–549. https://doi.org/10.1145/3025453.3025857

[17] Maria Christofi and Despina Michael-Grigoriou. 2018. Virtual reality for inducing empathy and reducing prejudice towards stigmatized groups: A survey. *Proceedings of the 2017 23rd International Conference on Virtual Systems and Multimedia, VSMM 2017* 2018-Janua (2018), 1–8. https://doi.org/10.1109/VSMM.2017.8346252

[18] Valve Corporation. 2022. *Steam VR - last accessed 01/05/2022*. https://www.steamvr.com/

[19] Dell. 2022. *Alienware Area51 Gaming Desktop - last accessed 01/05/2022*. https://www.dell.com/en-us/shop/gaming-and-games/alienware-area-51-gaming-desktop/spd/alienware-area51-r5

[20] Lina Eklund. 2011. Doing gender in cyberspace: The performance of gender by female World of Warcraft players. *Convergence* 17, 3 (2011), 323–342.

[21] Pedro Guillermo Feijóo-García, Mohan Zalake, Alexandre Gomes de Siqueira, Benjamin Lok, and Felix Hamza-Lup. 2021. Effects of Virtual Humans' Gender and Spoken Accent on Users' Perceptions of Expertise in Mental Wellness Conversations. In *Proceedings of the 21st ACM International Conference on Intelligent Virtual Agents* (Virtual Event, Japan) *(IVA '21)*. Association for Computing Machinery, New York, NY, USA, 68–75. https://doi.org/10.1145/3472306.3478367

[22] Dan Feng, David C. Jeong, Nicole C. Krämer, Lynn C. Miller, and Stacy Marsella. 2017. "Is It Just Me?": Evaluating Attribution of Negative Feedback as a Function of Virtual Instructor's Gender and Proxemics. In *Proceedings of the 16th Conference on Autonomous Agents and MultiAgent Systems* (São Paulo, Brazil) *(AAMAS '17)*. International Foundation for Autonomous Agents and Multiagent Systems, Richland, SC, 810–818.

[23] UNESCO Institute for Statistics. 2012. International standard classification of education: ISCED 2011. *Comparative Social Research* 30 (2012).

[24] Jesse Fox, Jeremy N Bailenson, and Liz Tricase. 2013. The embodiment of sexualized virtual selves: The Proteus effect and experiences of self-objectification via avatars. *Computers in Human Behavior* 29, 3 (2013), 930–938.

[25] Nirit Gavish, Teresa Gutierrez, Sabine Webel, and Jorge Rodriguez. 2011. Design Guidelines for the Development of Virtual Reality and Augmented Reality Training Systems for Maintenance and Assembly Tasks. *The International Conference SKILLS* 29 (2011), 1–4. https://doi.org/10.1051/bioconf/201101000

[26] VLimeSurvey GmbH. 2022. *LimeSurvey - last accessed 01/05/2022*. https://www.limesurvey.org

[27] Andrzej Grabowski and Jarosław Jankowski. 2015. Virtual Reality-based pilot training for underground coal miners. *Safety Science* 72 (2015), 310–314. https://doi.org/10.1016/j.ssci.2014.09.017

[28] Donna Jeanne Haraway. 1985. *A manifesto for cyborgs: Science, technology, and socialist feminism in the 1980s*. Center for Social Research and Education San Francisco, CA.

[29] Sandra G Harding. 1986. *The science question in feminism*. Cornell University Press.

[30] Susan Hawthorne and Renate Klein. 1999. *Cyberfeminism: Connectivity, critique and creativity*. Spinifex Press.

[31] Tanja Hentschel, Madeline E. Heilman, and Claudia V. Peus. 2019. The Multiple Dimensions of Gender Stereotypes: A Current Look at Men's and Women's Characterizations of Others and Themselves. *Frontiers in Psychology* 0, JAN (jan 2019), 11. https://doi.org/10.3389/FPSYG.2019.00011

[32] Laura Boteler Humm, Dale Olsen, Morris Be, Michael Fleming, and Matthew Smith. 2014. Simulated job interview improves skills for adults with serious mental illnesses. *Annual Review of CyberTherapy and Telemedicine* 12 (2014), 50–54.

[33] Emmanuel Johnson, Jonathan Gratch, Jill Boberg, David DeVault, Peter Kim, and Gale Lucas. 2021. Using Intelligent Agents to Examine Gender in Negotiations. In *Proceedings of the 21st ACM International Conference on Intelligent Virtual Agents* (Virtual Event, Japan) *(IVA '21)*. Association for Computing Machinery, New York, NY, USA, 90–97. https://doi.org/10.1145/3472306.3478348

[34] Sven Kachel, Melanie C. Steffens, and Claudia Niedlich. 2016. Traditional Masculinity and Femininity: Validation of a New Scale Assessing Gender Roles. *Frontiers in Psychology* 0, JUL (jul 2016), 956. https://doi.org/10.3389/FPSYG.2016.00956

[35] Alexandra Kitson, Mirjana Prpa, and Bernhard E. Riecke. 2018. Immersive interactive technologies for positive change: A scoping review and design considerations. *Frontiers in Psychology* 9, AUG (2018). https://doi.org/10.3389/fpsyg.2018.01354

[36] Gale M. Lucas, Nicole Krämer, Clara Peters, Lisa-Sophie Taesch, Johnathan Mell, and Jonathan Gratch. 2018. Effects of Perceived Agency and Message Tone in Responding to a Virtual Personal Trainer. In *Proceedings of the 18th International Conference on Intelligent Virtual Agents* (Sydney, NSW, Australia) *(IVA '18)*. Association for Computing Machinery, New York, NY, USA, 247–254. https://doi.org/10.1145/3267851.3267855

[37] Raquel Macedo, Nuno Correia, and Teresa Romão. 2019. Paralympic VR Game: Immersive Game Using Virtual Reality and Video. In *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems* (Glasgow, Scotland Uk) *(CHI EA '19)*. Association for Computing Machinery, New York, NY, USA, 1–6. https://doi.org/10.1145/3290607.3312938

[38] Joseph P McCleery, Ashley Zitter, Rita Solórzano, Sinan Turnacioglu, Judith S Miller, Vijay Ravindran, and Julia Parish-Morris. 2020. Safety and feasibility of an immersive virtual reality intervention program for teaching police interaction skills to adolescents and adults with autism. *Autism Research* 13, 8 (2020), 1418–1424.

[39] Robert R McCrae and Paul T Costa. 1987. Validation of the five-factor model of personality across instruments and observers. *Journal of personality and social psychology* 52, 1 (1987), 81.

[40] Microsoft. 2022. *Azure - Cognitive Services - Speech to Text - last accessed 01/05/2022*. https://azure.microsoft.com/en-us/services/cognitive-services/speech-to-text

[41] Procheta Nag and Özge Nilay Yalçın. 2020. Gender Stereotypes in Virtual Agents. In *Proceedings of the 20th ACM International Conference on Intelligent Virtual Agents* (Virtual Event, Scotland, UK) *(IVA '20)*. Association for Computing Machinery, New York, NY, USA, Article 41, 8 pages. https://doi.org/10.1145/3383652.3423876

[42] Dimitris Nathanael, Stergios Mosialos, and George-C. Vosniakos. 2016. Development and evaluation of a virtual training environment for on-line robot programming. *International Journal of Industrial Ergonomics* 53 (2016), 274–283. https://doi.org/10.1016/j.ergon.2016.02.004

[43] Laura P Naumann, Simine Vazire, Peter J Rentfrow, and Samuel D Gosling. 2009. Personality judgments based on physical appearance. *Personality and social psychology bulletin* 35, 12 (2009), 1661–1671.

[44] Quynh Nguyen, Emma Jaspaert, Markus Murtinger, Helmut Schrom-Feiertag, Sebastian Egger-Lampl, and Manfred Tscheligi. 2021. Stress Out: Translating Real-World Stressors into Audio-Visual Stress Cues in VR for Police Training. In *IFIP Conference on Human-Computer Interaction*. Springer, 551–561.

[45] Harley Ogier. 2020. The Player-Ethnographer: An Ethnographic Approach to the Study of Non-Player Characters in Digital Games. In *Proceedings of the Australasian Computer Science Week Multiconference* (Melbourne, VIC, Australia) *(ACSW '20)*. Association for Computing Machinery, New York, NY, USA, Article 43, 10 pages. https://doi.org/10.1145/3373017.3373064

[46] Daniel Pimentel and Sri Kalyanaraman. 2020. Your Own Worst Enemy: Implications of the Customization, and Destruction, of Non-Player Characters. In *Proceedings of the Annual Symposium on Computer-Human Interaction in Play* (Virtual Event, Canada) *(CHI PLAY '20)*. Association for Computing Machinery, New York, NY, USA, 93–106. https://doi.org/10.1145/3410404.3414269

[47] Sadie Plant. 1998. Coming across the future. *Virtual Futures* (1998), 39.

[48] Mariachiara Rapuano, Filomena Leonela Sbordone, Luigi Oreste Borrelli, Gennaro Ruggiero, and Tina Iachini. 2021. *The Effect of Facial Expressions on Interpersonal Space: A Gender Study in Immersive Virtual Reality*. Springer Singapore, Singapore, 477–486. https://doi.org/10.1007/978-981-15-5093-5_40

[49] Anne Reuten, Maureen van Dam, and Marnix Naber. 2018. Pupillary responses to robotic and human emotions: The uncanny valley and media equation confirmed. *Frontiers in Psychology* 9, MAY (may 2018). https://doi.org/10.3389/FPSYG.2018.00774

[50] Gabriel Rivera, Kenneth Hullett, and Jim Whitehead. 2012. Enemy NPC Design Patterns in Shooter Games. In *Proceedings of the First Workshop on Design Patterns in Games* (Raleigh, North Carolina, USA) *(DPG '12)*. Association for Computing Machinery, New York, NY, USA, Article 6, 8 pages. https://doi.org/10.1145/2427116.2427122

[51] Katja Rogers, Maria Aufheimer, Michael Weber, and Lennart E. Nacke. 2018. Exploring the Role of Non-Player Characters and Gender in Player Identification. In *Proceedings of the 2018 Annual Symposium on Computer-Human Interaction in Play Companion Extended Abstracts* (Melbourne, VIC, Australia) *(CHI PLAY '18 Extended Abstracts)*. Association for Computing Machinery, New York, NY, USA, 271–283. https://doi.org/10.1145/3270316.3273041

[52] Katja Rogers, Maria Aufheimer, Michael Weber, and Lennart E. Nacke. 2018. Towards the Visual Design of Non-Player Characters for Narrative Roles. In *Proceedings of the 44th Graphics Interface Conference* (Toronto, Canada) *(GI '18)*. Canadian Human-Computer Communications Society, Waterloo, CAN, 154–161. https://doi.org/10.20380/GI2018.21

[53] Helmut Schrom-feiertag, Florian Lorenz, Georg Regal, and Volker Settgast. 2018. Augmented and Virtual Reality Applied for Innovative , Inclusive and Efficient Participatory Planning. In *Proceedings of 7th Transport Research Arena TRA 2018*. 1–10.

[54] Leah Sheppard and Karl Aquino. 2013. Much Ado About Nothing? Observers' Problematization of Women's Same-Sex Conflict at Work. *Academy of Management Perspectives, The* 27 (02 2013), 52. https://doi.org/10.5465/amp.2012.0005

[55] Annika Silvervarg, Kristin Raukola, Magnus Haake, and Agneta Gulz. 2012. The Effect of Visual Gender on Abuse in Conversation with ECAs. In *Intelligent Virtual Agents*, Yukiko Nakano, Michael Neff, Ana Paiva, and Marilyn Walker (Eds.). Springer Berlin Heidelberg, Berlin, Heidelberg, 153–160.

[56] Christopher J. Soto and Oliver P. John. 2017. The next Big Five Inventory (BFI-2): Developing and assessing a hierarchical model with 15 facets to enhance bandwidth, fidelity, and predictive power. *Journal of Personality and Social Psychology* 113, 1 (jul 2017), 117–143. https://doi.org/10.1037/PSPP0000096

[57] Unity. 2022. *Real Time Development Platform - last accessed 01/05/2022*. https://unity.com/

[58] Eva A Van Reijmersdal, Jeroen Jansz, Oscar Peters, and Guda Van Noort. 2013. Why girls go pink: Game character identification and game-players' motivations. *Computers in Human Behavior* 29, 6 (2013), 2640–2649.

[59] VIVE. 2022. *Vive Pro Eye - last accessed 01/05/2022*. https://www.vive.com/eu/product/vive-pro-eye/overview/

[60] Judy Wajcman. 2006. Technocapitalism meets technofeminism: women and technology in a wireless world. *Labour & Industry: a journal of the social and economic relations of work* 16, 3 (2006), 7–20.

[61] Judy Wajcman. 2010. Feminist theories of technology. *Cambridge journal of economics* 34, 1 (2010), 143–152.

[62] Candace West and Don H Zimmerman. 1987. Doing gender. *Gender & society* 1, 2 (1987), 125–151.

[63] Nick Yee and Jeremy Bailenson. 2007. The Proteus effect: The effect of transformed self-representation on behavior. *Human communication research* 33, 3 (2007), 271–290.

[64] Nick Yee, Nicolas Ducheneaut, Mike Yao, and Les Nelson. 2011. Do men heal more when in drag? Conflicting identity cues between user and avatar. In *Proceedings of the SIGCHI conference on Human factors in computing systems*. 773–776.

[65] Katja Zibrek, Benjamin Niay, Anne-Hélène Olivier, Ludovic Hoyet, Julien Pettre, and Rachel McDonnell. 2020. The Effect of Gender and Attractiveness of Motion on Proximity in Virtual Reality. *ACM Trans. Appl. Percept.* 17, 4, Article 14 (nov 2020), 15 pages. https://doi.org/10.1145/3419985