



Insights from the Uncanny Valley: Gender(Sex) Differences in Avatar Realism and Uncanniness Perceptions

Jacqueline D. Bailey^(✉) , Karen L. Blackmore , and Robert King

College of Engineering, Science and Environment, University of Newcastle, Ring Road,
Callaghan, NSW, Australia

{jacqueline.d.bailey, karen.blackmore,
robert.king}@newcastle.edu.au

Abstract. Two core factors influence the perception of avatars. On one side are the developers who are concerned with building avatars and in some cases pushing the boundaries of the realism. These developers are constrained by the resources available to them that allow them to produce the optimal avatar with the equipment, time, and skills available to them. On the other side are users who engage with avatars who are directly affected by the choices the developers have made. Inside the interaction between these sides is the avatar itself, whose appearance, level of realism and fundamental characteristics like a perceived gender(sex) can influence the user's perception of that avatar. Despite the large amount of research dedicated to understanding the perception of avatars, many gaps in our understanding remain. In this work, we aim to contribute to further understanding one of these gaps by investigating the potential role of gender(sex) in the perception of avatar realism and uncanniness. We add to this discussion by presenting the results of an experiment where we evaluated realism and uncanniness perceptions by presenting a set of avatars to participants ($n = 2065$). These avatars are representative of those used in simulation and training contexts, from publicly available sources, and have varying levels of realism. Participants were asked to rank these avatars in terms of their realism and uncanniness perceptions to determine whether the gender(sex) of the participant influences in these perceptions. Our findings show that the gender(sex) of a participant does affect the perception of an avatar's realism and uncanniness levels.

Keywords: Uncanny valley · Uncanniness · Avatar · Human computer interaction · Gender(sex)

1 Introduction

Avatars have diverse appearances and can be used to represent a person or characters in mixed reality, online forms, gaming, training, and simulation contexts. Regardless of the intended purpose of an avatar, at its core they function as visual representations of characters used in a specific setting. Commonly, avatars (virtual humans) are used in

serious games and simulation training scenarios where they can engage with an end-user for training purposes [1, 2]. A core issue for all avatars regardless of their role is the end-users' perception of that avatar. Much of the literature on avatar perceptions fall into two core categories, focusing on physical attributes like hair or eye color or investigating how more functional aspects such as perceived realism and uncanniness levels can affect the perceptions of avatars.

The focus of this work is an examination of gender(sex) in the realism and uncanniness perceptions of avatar faces. Following Stumpf, Peters [3], this work draws on the socially constructed perceptions of gender, here referred to as gender(sex). Whereby gender identification, gender expression and performance might not necessarily align with biological sex of the end-user. The perceived gender(sex) of an avatar may influence the user's perception of the avatars ability which may be attributed to associated stereotypical ideas of gender. Further, the level of sexualization in the physical appearance of an avatar can also affect their perceived abilities [4]. It is also worth noting that some of the content of this chapter was previously presented at the International Conference on Computer-Human Interaction Research and Applications (CHRIA22) [5]. The current work has been extended to include additional content from the survey described in the original paper. Also included is an updated literature review, and discussion of the key findings for this new work.

Focusing on this current work it is important to discuss how the perception of avatars can be influence by developmental characteristics. For example, resources such as time, finances, tools, and expertise used to create an avatar can directly affect the level of realism that is realistically achievable given the constraints. These resources can range widely, from high-end tools like a light stage which surround an actor in a spherical structure of lights that can capture data such as specular mapping, facial geometry and surface reflections which can be used to create a highly realistic avatar [6] to produce higher realism avatars. By comparison, using commercial off the shelf software may have a lower level of realism due to the limited resources available [7]. While not as 'high-tech', this option is a relatively accessible set of tools for developers with limited expertise and/or finances.

Realism has many facets and can refer to an avatar's behavior, appearance, or ability to communicate [8]. The two main areas that may affect this perception are the visual and behavioral nuances shown. This includes how these visual aspects relate to the perceived anthropomorphism of an avatar and the kinetic similarity and social appropriateness of an avatars' behavior. These aspects, individually or combined, can define how some end-users perceive realism in avatars [9].

When an avatar fails to meet the visual, kinetic, or behavioral fidelity of a healthy human they can be perceived as uncanny or eerie. This uncanniness can be linked to the Mori's Uncanny Valley Theory [10] in which the human likeness perceptions experience a sharp dip into a negative familiarity when these expectations are not met, as can be seen in Fig. 1.

Mori suggests that this instinctive response can be linked to a form of protection for the viewer, that protects them from proximal sources of dangers. For example, within the negative familiarity dip we find unpleasant or potentially harmful sources of anxiety for human viewers such as corpses. Although this theory was originally applied to the field

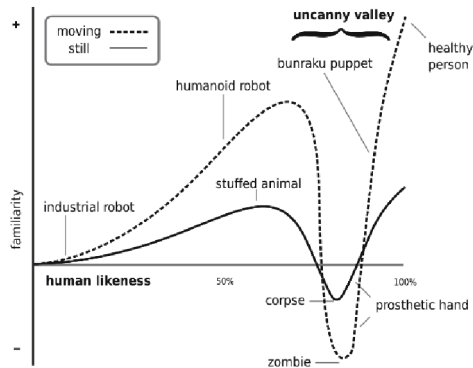


Fig. 1. The Uncanny Valley [10].

of robotics it has now been repeatedly applied to human-like avatars [11–13]. These uncanniness perceptions might be the result of creating an avatar using sophisticated software that focuses on boundary pushing visual realism levels which may trigger the uncanniness or eerie perceptions in end-users [11].

The next section provides a comprehensive overview of the key literature relating to the importance of visual realism in avatars and consequences of this realism falling short of expectations. Also discussed is the influence of gender(sex) in the perceptions of avatars. Then we describe the methodology for our avatar ranking survey including data collection techniques and procedure before presenting the key results of the data analysis. Our paper concludes with a discussion of the key findings alongside some recommendations for future work in this area.

2 Background

There are several important areas to consider in the design and development of avatars. One overriding area that can have a deep effect on the perception of an avatar is the perceived level of visual realism. Avatars are often generated with the resources available to the designers and developers, within the constraints of time, money, expertise, and resources, with the resulting avatar being the optimal output based on these constraints. However, the importance of visual realism can lead to an unintended negative perception in terms of how uncanny or eerie an avatar appears. While the avatars are constrained by their designers and developers' resources, the consequences of uncanniness cannot be ignored. The uncanniness perceptions, as described in the following section, can lead to users rejecting an avatar, which in turn leads to wasted resources used to develop and design that avatar. The perception of avatars can be influenced by many factors, but arguably one of the core design choices that developers make, is the gender(sex) of an avatar. As discussed below, this choice can be a very deliberate one that nurtures expectations or stereotypes, or it can be unintentional. Regardless of the design decisions the literature has shown that the influence of gender(sex) is important. The importance of gender(sex) perceptions also extends to the perception of realism levels in avatars.

2.1 The Importance of Visual Realism in Avatars

The realism level of an avatar is prevailing issue that impacts the development and appearance of an avatar. The level of realism can have a profound influence on the perception of that avatar. A study by Dobre, Wilczkowiak [14] shows that photo-realistic avatars were seen as more trustworthy and were assigned higher affinity scores. This level of trust can be important in human-avatar interactions especially if an avatar is being used in serious purposes such as medicinal [1] or military [15] simulation training scenarios.

Zanbaka, Goolkasian [13] have also examined the influence of an avatar's realism and how this may affect their persuasiveness. Zanbaka et. al's study used real human actors, human-like avatars, and anthropomorphized cat avatars to measure persuasiveness. Their findings suggest that the realism levels did not impact the persuasiveness of an avatar's message. As previously stated, avatars are a product of purposeful design, and the resources available to produce them can only create an avatar that is optimal for the circumstance. This purposeful design influences the avatar design issues in serious game applications may be less about financial loss, and more about poor assistance, learning, and training experiences. This impact may occur when either the behavioral or visual realism of an avatar negatively affects the intended end-user, through uncanniness perceptions [16].

Consistent with existing literature, the realism level has a direct impact on the perception of uncanniness in avatars. For example, MacDorman, Green [17] suggest that end-users are disturbed when a virtual character's (avatars) appearance is too realistic or human. Tinwell [18] suggests that increasing the level of realism does not necessarily mean that the acceptance of the avatar will increase. Thus, these trade-offs between avatar realism the effects of uncanniness perceptions can be complex with serious consequences when the avatar fails to meet end-user expectations.

2.2 When Visual Realism Fails, the Consequences Are Uncanny and Eerie

As mentioned previously, the feeling of uncanniness that end-users may experience when engaging with boundary pushing avatars can be a major handicap for human-avatar interaction. The consequences of this boundary pushing can lead to the avatars falling into the previously mentioned Uncanny Valley [10]. These consequences, while present may still allow avatars to be effective as discussed by Yoon, Kim [19] who used wooden mannequins and high-realism avatars in their study and found that there was a strong acceptance of sense of realism from the high realism avatars despite the participants experiencing the negative effects of the uncanny valley.

These judgements can be based on what Ambady, Bernieri [20] call thin slices, whereby someone who perceives another person can form relatively quick and accurate social judgements of others with only the minimal amount of information. This is an important distinction to understand as making accurate social judgements about other persons is a fundamental human trait for forming successful relations and avoiding potentially harmful interactions [21]. This avoidance of harm or danger is echoed by MacDorman, Green [17] who suggest that the uncanniness feeling that end-users feel

can also be associated with threat avoidance. If an avatar provokes such a reaction in the end-user, when their intended purpose was to help the user, the avatar will be rejected.

Such a rejection is a serious consequence of uncanny or eerie avatars and can lead to serious financial losses for companies. For example, financial losses were evident for Disney in the movie 'Mars Needs Moms' which purportedly cost the Disney Corporation \$150 million [22] when the movie flopped. This can also be seen in games such as *L.A. Noire* and *Mass Effect: Andromeda* where players have heavily criticized the appearance of the in-game avatar [22]. These examples of uncanny virtual characters can also be attributed to the human visual system.

This visual system can easily and quickly detect falsehoods found in human-like faces based on previous exposure to human faces [12]. Avatars that attempt to reflect an accurate representation of a human face are susceptible to the effects of the uncanny valley. This uncanniness is exacerbated when the avatar moves or attempts to express emotions in realistic ways. Several methods have been developed to generate these emotions with varying levels of success.

Methods for capturing and expressing emotions can range from manual methods such as frame-by-frame approach, to motion-capture to more advanced techniques that use machine learning. The manual methods may effectively create facial animation, but it is an extremely time-consuming and expensive process. Alternative techniques such as motion capture do exist and can provide a somewhat faster means of capturing an actor's facial movement. When captured either as a real-time motion capture or as a set of motions mapped post-capture [23] can create effective expressions for avatars. Alternatively, there are machine learning techniques which use neural rendering [24–26], which can render highly photorealistic avatars. However, despite these differences, very little is understood about the impacts of these different techniques on perceptions of realism or uncanniness and how an avatar's gender(sex) may influence these perceptions.

2.3 The Influence of an Avatar's Gender(Sex)

A fundamental design decision when creating avatars is the choice of their perceived gender(sex). Existing research explores several gender(sex) related issues including gender(sex) stereotypes, the proteus effect, and gender(sex) swapping. Fox and Bailenson [27] observed some gender(sex) based stereotypes and suggest that female avatars are more likely to appear either as hypersexualized or as an ornament within video/computer games or as a support role. The perceived level of sexualization that an avatar is depicted as having can also affect female avatars' perceived abilities [4]. The Proteus Effect as discussed by Yee and Bailenson [28] suggests that the perceived gender(sex) of an avatar will lead users to conform to behavioral expectations they have associated with a specific gender(sex). This was also observed by Beltrán [29] who suggested that this conformity exists in simulation and training contexts as well. Their findings suggest that using a male avatar to train professional women will negatively affect her and her colleagues' achievements. This is essential to consider, as Beltrán [29] argues that most simulation tools show a generic male avatar during training. This raises an important issue to consider for avatar designers and developers, if an avatar is designed for a specific purpose, it is worth considering who in the user-base the avatar is likely to encounter. For example,

if an avatar is intended to train a largely female orientated cohort, a generic male avatar may not be the best choice.

Other research in the influence of gender(sex) investigates gender(sex) swapping and exploring an end-user's gender identity or identities. Lehdonvirta, Nagashima [30] also suggest that male participants are more likely to seek and receive help when they are 'disguised' as a female-styled avatar. This was also investigated by Hussain and Griffiths [31] who suggest that there are many social benefits to gender(sex)-swapping in online gaming. For example, male players engaging as a female-styled avatars may do so to be more favorably treated by other male players to gain benefits or favor within the context of the video/computer game.

Lastly, the gender(sex) can also be used to explore and express gender identity or identities [32]. This is an important aspect to consider as it allows users to express their gender identities which may or may not reflect the gender identity they present in their everyday lives. In video/computer gaming settings, players may be able to remain relatively anonymous online while they explore and express their gender identity or identities in a reasonably safe platform.

Thus, based on the existing literature, we have completed a ranking exercise with a set of ten homogenous avatars to better understand how perceptual variability may be influenced by gender(sex). This will also create a robustly ranked set of avatars for use in future research. These avatars will have been quantifiability rated in terms of realism and uncanniness perceptions. The methodology for data collection is discussed next in this work. While the set of avatars used in our study are of similar age and ethnicity, they have distinctly varying levels of realism and are from multiple sources. In summary, we investigate the relationship between a set of avatar faces realism and uncanniness perceptions with a focus on how gender(sex) may affect these perceptions. To examine the potential influence of gender(sex) in the perception of avatar realism and uncanniness, we conducted a mass-scale survey, described next.

3 Avatar Ranking Survey Methodology











Our study uses a set of ten avatars to assess the potential influence of gender(sex) in the perception of realism and uncanniness. The survey was produced in Limesurvey [33] and hosted by Amazon's Mechanical Turk (Mturk) [34] and took participants 15–20min to complete. The participants who met the inclusion criteria were paid .10USD cents for their participation. Our study has been approved by the University of Newcastle's Ethics Committee (Protocol number: H-2015–0163).

With a mean age of 34.82 years a total of ($n = 2065$) participants completed the Avatar Ranking Survey. To examine potential differences between genders(sexes) participants were asked to nominate both a biological sex, and a gender identity. A small percentage of the participants' gender identities did not necessarily align with their nominated biological sex, but as this number was small, we opted to examine gender(sex) by the biological sex indicated by the participants. In total, there were 1050 self-identified female participants, 1003 male, three people self-identified as transgender, two selected 'other' as their gender and seven people chose not to say which gender(sex) they are.

3.1 Sample Avatar Set Images

Our 10 homogenous avatar faces (Table 1) are a sample that broadly captures varying levels of realism which can be achieved from different creation methods discussed previously. These avatars only appear to represent assumed binary genders (sexes) with five female faces and five male faces. The set also includes two real human faces sourced from an online dataset (DaFEx [35]). This set of avatars are representative of avatars found in training and simulation contexts and are from several sources [35–40].

Table 1. Sample images of the avatar faces [5].

Female Avatar Faces				
				
Rose	Emily	Ilana	Liliwen	Bailie
Male Avatar Faces				
				
Rycroft	Ira	Victor	Macaw	Leo

3.2 Avatar Ranking Survey Procedure

Our approach is similar to Lange [41] who used a ranking exercise to examine images of virtual landscapes. To ensure a higher reliability a ranking approach is used as opposed to a rating method in both this study and Lange’s [42, 43]. Participants were asked to rank the avatars twice. First from most to least realistic and second from most to least uncanny or eerie (see Fig. 2).

Prior to ranking the avatars, participants were asked a series of demographic questions that asked them to indicate a biological sex, gender identity, age, English language proficiency, and Country of residence. Additional questions were asked to determine each participant’s level of computer/video gaming, virtual environment, and avatar animation experience to potentially give further context to the ranks when analyzing the data.

3.3 Data Analysis for the Survey Responses

Both the realism and uncanniness ranks were first analyzed using Friedman tests. If the initial test returned a statistically significant result, a post hoc Wilcoxon signed-rank test

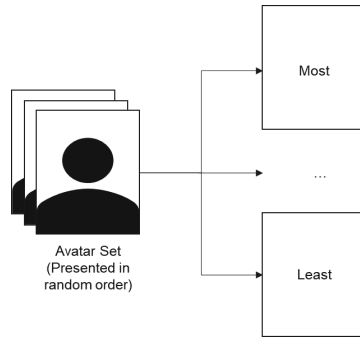


Fig. 2. Avatar Ranking Survey – ranking questions.

was performed. The second test was run with a Bonferroni adjustment for multiple tests were run to investigate any differences between the rankings.

Primarily, the participant gender(sex) variable was investigated to determine whether this factor influenced the realism and uncanniness rankings. Our approach was based on the work of Conover and Iman [44] who suggest that rank transformation procedures should allow the use of parametric methods. But, as the participants themselves provided the ranks the transformation step was not needed.

Based on the Friedman tests, additional analysis of the uncanniness ranks was run using a General Linear Model (GLM). The effect of the participants gender(sex) was examined using a full factor two-way ANOVA model with avatar and gender(sex) being the two core factors. This was followed up with a pairwise comparison to determine whether there were significant differences using a Bonferroni adjustment to control for the familywise error rate of multiple tests, the results of this analyses are presented in the following section.

4 Results from the Avatar Ranking Survey











This section focuses on the two sets of ranking data by first examining the realism ranks and the potential influence of participant gender(sex). Second, we outline the uncanniness ranks and how gender(sex) may impact on the participants perceptions.

4.1 Realism Ranks and Gender(Sex)

Table 2 below shows the mean of the realism rankings for each avatar. Based on the Freidman test, there was a statistically significant difference in the perception of the avatars' realism $\chi^2(9) = 8819.9, p < .001$.

The human actors are considered the most realistic (Rose ($M = 2.67, SD 2.199$) and (Rycroft ($M = 3.06, SD = 2.147$)). While it is an expected result, it is noted that the participants were not told that some of the avatar faces were in fact real world humans; this somewhat validates the efficacy of the ranking set. Next, the avatars created by the University of Southern California and Image Metrics are ranked as the third and fourth

Table 2. Overall Ranking of the avatars by their perceived level of Realism.

Rank	Avatar	Name	Realism Level	Gender(Sex)	Mean	SD
1		Rose	Real	Female	2.67	2.199
2		Rycroft	Real	Male	3.06	2.147
3		Ira	High	Male	3.53	2.020
4		Emily	High	Female	3.57	2.085
5		Victor	Mid1	Male	6.06	1.915
6		Ilana	Mid1	Female	6.12	1.938
7		Macaw	Mid2-Low	Male	7.07	1.919
8		Victor	Mid2-Low	Male	7.08	2.027
9		Liliwen	Mid2-Low	Female	7.56	2.404
10		Bailie	Mid2-Low	Female	8.27	2.122

highest realism avatars in this set (Ira ($M = 3.53$, $SD = 2.020$) and Emily ($M = 3.57$, $SD = 2.085$) behind the real-world human faces. Then the avatars sourced from Faceware [38] are rated as the fifth and sixth most realistic of the set (Victor ($M = 6.06$, $SD = 1.915$) and Ilana ($M = 6.12$, $SD = 1.938$)). The FaceShift [37] avatars are ranked seventh (Macaw ($M = 7.07$, $SD = 1.919$)) and eighth (Leo ($M = 7.08$, $SD = 2.027$)) for their perceived level of realism. Importantly, we see that the lower realism ranks are both females, with the lower realism avatars are Liliwen ($M = 7.56$, $SD = 2.404$) and Bailie ($M = 8.27$, $SD = 2.122$)).

A Wilcoxon signed ranked test using a Bonferroni correction determined that there are no statistically significant differences between the following comparisons of pairs; both of the Mid2-Low realism male avatars Leo and Macaw, the high realism avatars Ira and Emily and finally, the Mid1 realism avatars Victor and Ilana, in terms of realism. All other comparisons were statistically significantly different.

We first consider potential gender(sex) impacts through an examination of the realism ranking scores by participant sex. Using a General Linear Model, there was a statistically significant interaction between the avatars and the participants' gender(sex) $F(9,19910) = 16.34, p < .001$. Posthoc analysis used .005 as the significance level (Bonferroni adjustment for ten tests) to compare differences between the responses of female and male participants for each avatar. Female participants rated Rycroft the Real Human Male as more realistic than their male counterparts: (Female ($M = 2.85, SD = .068$), Male ($M = 3.30, SD = .064$), $p < .001$). In contrast, Rose the Real Human Female was rated as more realistic by female participants when compared to the male participants. (Female ($M = 2.74, SD = .068$), Male ($M = 2.93, SD = .064$), $p < .001$). The overall ranks for the uncanniness scores are discussed in the following section.






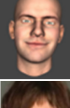


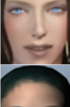

4.2 Uncanniness Ranks and Gender(Sex)

The Friedman test for the uncanniness rank scores revealed that there is a statistically significant difference in the avatar sets uncanniness perceptions $\chi^2(9) = 156.254, p < .001$. It is noteworthy, that the mean scores for the uncanniness ranks are clustered between 4.9–5.8, which may suggest that variations in uncanniness perceptions may be subtle (Table 3).

Despite being from the same creation tool and being subjected to the same creation method, we see that the Mid1 realism avatars are polar opposites within these uncanniness ranks. With Victor the Mid1 realism male avatar being ranked as the uncanniest avatar in the set ($M = 4.99, SD = 2.505$), while in contrast Ilana (Mid1 realism female) has been ranked as the least uncanny avatar in the set ($M = 5.87, SD = 2.324$). This trend of the male avatars being uncannier than their female counterparts continues for the high realism avatars with Ira the high realism male ($M = 5.14, SD = 2.805$) being considered uncannier than Emily ($M = 5.37, SD = 3.006$). This trend is similar for one of the Mid2-Low male avatars (Leo ($M = 5.47, SD = 2.739$)) and Rycroft the real human male ($M = 5.51, SD = 3.114$). Four out of five of the upper ranks appear to be populated by male faces which may suggest some gendered affect in the perception of uncanniness in avatar faces. The only exception to this is Emily the high realism female avatar.

We see this trend is reversed for the bottom five ranks with Macaw the other Mid2-Low realism male avatar being the only male avatar in the lower ranks ($M = 5.52, SD = 2.571$). The bottom four ranks are populated exclusively by female avatars, which may suggest some gendered affect to the perception of uncanniness. Like other avatars, Rose the real human is ranked as less uncanny than her male counterpart ($M = 5.65, SD = 3.360$). However, it is noteworthy that the participants were not told there may be real faces in the avatar set, and as outlined above. Rose and Rycroft were ranked as the most realistic, but here they are ranked roughly around the middle of the uncanniness ranks. This middle ranking may suggest that the participants were uncertain as to whether these were 'real humans' or avatars. The literature and realism rankings above may suggest







Table 3. Overall Ranking of the avatars by their perceived level of Uncanniness [5].

Rank	Avatar	Name	Realism Level	Gender(Sex)	Mean	SD
1		Victor	Mid1	Male	4.99	2.505
2		Ira	High	Male	5.14	2.805
3		Emily	High	Female	5.37	3.006
4		Leo	Mid2-Low	Male	5.47	2.739
5		Rycroft	Real	Male	5.51	3.114
6		Macaw	Mid2-Low	Male	5.52	2.571
7		Rose	Real	Female	5.65	3.360
8		Bailie	Mid2-Low	Female	5.66	3.214
9		Liliwen	Mid2-Low	Female	5.83	2.799
10		Ilana	Mid1	Female	5.87	2.324

that as Rose and Rycroft were ranked as “most realistic” or possibly have been identified as real, that they should be ranked as “most uncanny” (Rank 1 or 2) but as the analysis shows, this is not the case suggesting that some other variable may have been at work in the participants decision making process.

We also see that despite Bailie one of the Mid2-Low realism female avatars being ranked as the least realistic, she was not ranked as the uncanniest avatar ($M = 5.66$, $SD = 3.214$), which may suggest that her appearance is somewhat cartoonish or has a level of mid-low realism that could be seen as uncanny to the participants.

Table 4. Summary of the gender(sex)-based variations in the uncanniness ranks from the GLM analysis.

Avatar	Category	Female Participants (M, SD)	Male Participants (M, SD)
	Real human female	(M= 5.85, SD= .093)	(M= 5.33, SD= .088)
	Real human male	(M= 5.68, SD= .093)	(M= 5.26, SD= .088)
	High Realism Female	(M= 5.54, SD= .093)	(M= 5.12, SD= .088)
	High Realism Male	(M= 5.27, SD= .093)	(M= 4.90, SD= .088)
	Mid2-Low Realism Female	(M= 5.47, SD= .093)	(M= 5.98, SD= .088)
	Mid2-Low Realism Male	(M= 5.30, SD= .093)	(M= 5.75, SD= .088)

To determine where the differences occurred in the uncanniness ranks a Wilcoxon signed ranked test using a Bonferroni correction was used. Despite the small range of scores there are some significant differences between some of the avatars ranks. Specifically, between Ira the high realism male and Rycroft the real human male ($Z = -.527, p. < 0.001$) and Ira the high realism male and Rose the real human female ($Z = -6.229, p. < 0.001$). A potential effect of uncanniness perceptions may be seen in a comparison of Emily the high realism female avatar who, unlike her counterpart, is

not statistically significant when compared to the real humans. This may suggest that gender(sex) may influence the perception of uncanniness when compared to the real humans in the set.

4.3 The Influence of Gender(Sex) in the Uncanniness Ranks

After determining the ranking of most to least uncanny avatars in the set we consider whether there were any gender(sex)-based variabilities amongst the scores. Using a GLM with a Bonferroni correction we analyzed the ranks and determined that there was a statistically significant interaction between the avatar set's uncanniness rank scores and the gender(sex) of the participants $F(9, 19910) = 10.453, p < .001$.

When examining the Posthoc analysis several significant results showed that there were some differences between the rank scores of the female and male participants (Table 4).

From this analysis we can see that most of the significant differences have the female participants rating the avatars as uncannier than their male counterparts. This can be seen in the analysis of the real humans, high realism avatars and one of the Mid2-Low realism male avatars. As can be seen in Table 4 Rose the human female (Rose $p < .001$, (Female (M = 5.85, SD = .093), Male (M = 5.33, SD = .088))), Rycroft the human male (Rycroft $p = .001$, (Female (M = 5.68, SD = .093), Male (M = 5.26, SD = .088))), are both scored as more uncanny by the female participants. This trend continues for both Emily the high realism female avatar (Emily $p = .001$, (Female (M = 5.54, SD = .093), Male (M = 5.12, SD = .088))), and Ira the high realism male avatar (Ira $p = .005$, (Female (M = 5.27, SD = .093), Male (M = 4.90, SD = .088))). We also see this dominance of female participants finding some avatars uncannier than their male counterparts occur for Macaw one of the Mid2-Low realism male avatars (Macaw $p = .001$, (Female (M = 5.30, SD = .093), Male (M = 5.75, SD = .088))).

The only instance where the male participants find an avatar uncannier than their female counterparts is for Bailie one of the Mid2-Low realism female avatars (Bailie $p < .001$, (Female (M = 5.47, SD = .093), Male (M = 5.98, SD = .088))). These differences suggest that the gender(sex) of the participant may influence their perceptions of an avatar's uncanniness.

4.4 Uncanniness Ranks Distribution

As previously mentioned, the actual scores for the rank scores range between 4.5–5.8 which may suggest that uncanniness perceptions can be subtle. To further investigate this, we examined the distributions of the raw uncanniness rank scores grouped by participant gender(sex) (Fig. 3) where we see several interesting patterns emerge.

It is noteworthy that the distributions themselves have appear to have a pattern. We can observe that most avatars have a flat uniform distribution for their scores. However, the Mid1 realism avatars (Victor and Ilana) and Mid2-Low realism male avatar (Macaw) appear to have a normal distribution of the rank scores. But interestingly, the real humans (Rose and Rycroft) and one of the Mid2-Low realism female avatars (Bailie) have a distinct bi-modal distribution showing the divided opinion of these specific avatars in terms of the participants uncanniness perceptions.

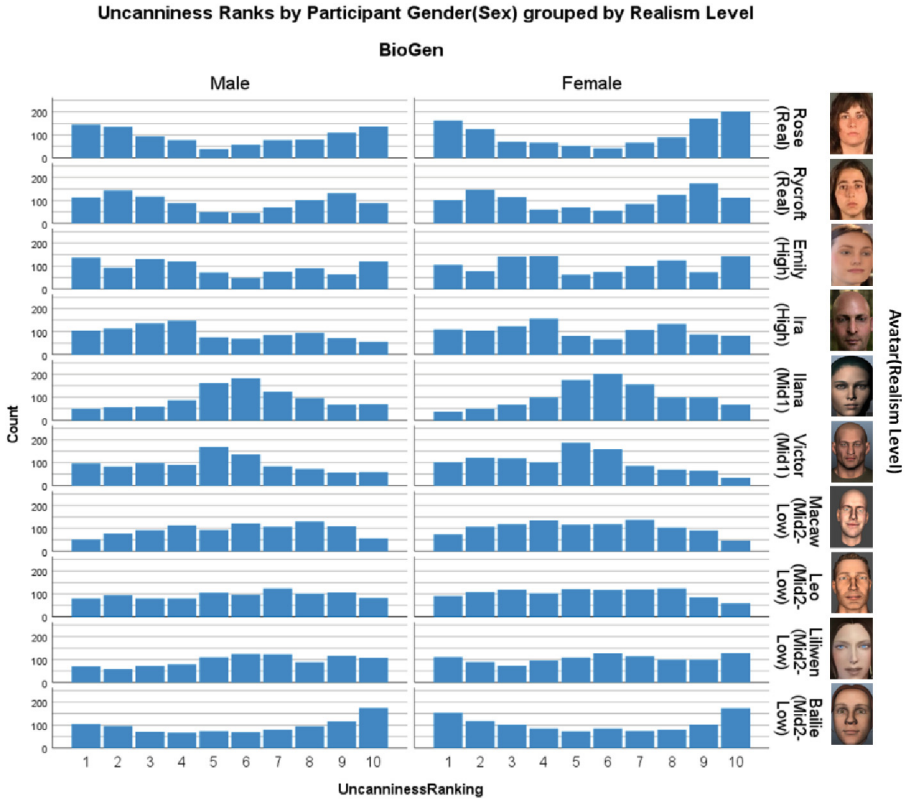


Fig. 3. Rank scores comparison by avatar and participant gender(sex) [5]

To determine whether these distributions are significant, we used a 2 sample Kolmogorov-Smirnov Test in order to compare the distribution shapes, using a Bonferoni correction for testing multiple pairs. Some significant differences were found when comparing the male and female distributions. First, we see these differences between Rose $D(1993) = 1.785, p. < .001$, Rycroft $D(1993) = 1.819, p. = .003$, Macaw $D(1993) = 1.796, p. = .003$, and Baillie $D(1993) = 1.785, p. = .003$.

Second, we also see some significant differences between Leo a Mid2-Low realism male avatar and Liliwen a Mid2-Low realism female avatar. However, these significant differences are only present for the female participants scores not the male participants (Female participants($D(2090) = 1.947, p. < .001$., Male participants($D(1896) = 1.355, p. = .051$)). This trend continues for Rose the real human and Emily the high realism female avatar (Female participants($D(2090) = 3.412, p. < .001$.), Male participants($D(1896) = 1.447, p. = .030$)). We see the trend reflected in the scores for Rycroft the real human male and Rose the real human female (Female participants($D(2090) = 1.312, p. = .064$), Male participants($D(1896) = 3.834, p. < .001$)). Finally, the female participants scores for both the Mid2-Low realism female avatars Liliwen and Baillie, show some significant differences where the male participants do not (Female

participants($D(2090) = 2.166, p. < .001.$, Male participants($D(1896) = 1.699, p. = .051.$)).

Third and interestingly, we see this trend reversed for some of the avatars in the set. Specifically, for Macaw one of the male Mid2-Low realism avatars and Ira the high realism male, the male participants scores are significant while the female participants are not (Male participants($D(1896) = 3.834, p. < .001$), Female participants($D(2090) = 1.312, p. = .064$)). This reversed trend continues with Bailie one of the Mid2-Low realism female avatars when compared to Ira the high realism male avatar, the male participants scores are statistically significant, whereas the female participants scores are not (Male participants($D(1896) = 2.825, p. < .001$), Female participants($D(2090) = 1.444, p. = .031$)).

Further, the comparison between Leo one of the Mid2-Low realism male avatars and Ira the high realism male were statistically significant for the male participants but not for the female participants (Male participants($D(1896) = 3.789, p. < .001$), Female participants($D(2090) = 1.553, p. = .016$)). Lastly, we see that this trend continues for a comparison between Emily the high realism female and Leo one of the Mid2-Low realism male avatars (Male participants($D(1896) = 3.330, p. < .001$), Female participants($D(2090) = 1.837, p. = .016$)). All other comparisons were not statistically significant for either participants gender (sex). The key findings of this analysis are discussed next.

5 Study Discussion and Conclusions

Our initial findings considered realism perceptions and show an interesting result in the grouping of the avatars by their source or creation method. It is unsurprising that avatars like Emily the high realism female and Ira the high realism male avatar would rank higher than other avatars due to the resources used to create them. While in contrast, Bailie one of the Mid2-Low realism female avatars which was created using an off the shelf model from the 3D Avatar store (<https://www.3d-avatar-store.com/>), a laptop and a GoPro, and expectedly achieved the lowest rank. The avatar realism rankings do appear to form natural groupings that align to the underlying method of creation. However, there are other factors likely to impact on the rankings. For instance, the facial avatars presented in this study are not consistent in terms of their features. End-user perceptions of falsehoods in the avatar faces overall appearance may have influenced realism perceptions as suggested by Brenton, Gillies [45]. This may lead to those with unusual characteristics being ranked as the least realistic avatars, however, this was not explicitly considered in this research. As expected, participants rated the human actors at the highest levels of realism, confirming the validity of the realism ranking process. Participants were able to correctly rank these faces despite not being identified as non-avatar.

It is interesting to note that although the participants were not told there would be human actors in the set, both Rose the real human female and Rycroft the real human male sit around the middle of the uncanniness rankings. Unlike the realism rankings that place the actors as highly realistic, the ambiguity over whether they are an avatar may have caused them to appear as mildly uncanny to the viewers. Jentsch [46] suggests

that a predominant cause of uncanniness is the doubt over the living status of an entity. This uncertainty over whether these ‘avatars’ are real may have contributed to them being ranked as mildly uncanny. In contrast, it is unsurprising that the higher realism avatars (Emily and Ira) were ranked as profoundly uncanny, as it has been suggested that increasing the level of realism in the stimuli presented would likely raise the level of sensitivity to cues that would indicate falsehood [45]. In addition to these general findings, two specific investigations were conducted into differences in realism and uncanniness perceptions based on the different genders (sexes) of participants.

The first investigation into the impact of gender (sex) on the rankings examined whether an observer’s gender (sex) affected the perception of an avatar’s realism ratings. The results suggest that the gender (sex) of a participant may affect the rating of an avatars’ realism level for avatars who may be perceived as mid-realism. There was some consensus between female and male participants at the extreme ends of the realism ratings. However, female participants rated Ira the high realism male avatar higher than Emily the high realism female avatar, while male participants had the reverse. At the other extreme, for the lowest-rated avatars, both female and male participants rated both the Mid2-Low realism female avatars Liliwen and Bailie as the least realistic.

Although the grouping by creation method continued in the 5th-8th realism ranks, the ordering was different for female and male participants. Female participants rated female avatars as more realistic than male avatars, whereas male participants did the reverse. This may be explained by the similarity-attraction theory that predicts preferences for gender (sex)-matching, which suggests that some females may see males as the more dominant group, and therefore use similarities to develop social bonds with other females [64]. However, as the female participants ranked Emily the higher realism female avatar lower than Ira the high realism male avatar, this similarity-attraction theory may not apply to all avatars. The impact of a participants’ self-similarity with each of these avatars may also influence these rankings. This data was collected as part of this ranking exercise and will be fully explored in future work.

Our investigation into the potential influence of gender (sex) also considered whether a participants’ gender (sex) affected perceptions of an avatar’s uncanniness level. We found that the Male participants appeared to rate the higher realism avatars as uncannier than all other avatars, with Victor the Mid1 realism male avatar being the exception. This is similar to previous findings for male participants where existing literature suggested that they were more sensitive to uncanniness in human-like avatars [11].

Of interest, are the top four ranks for female participants, these were populated exclusively by male avatar faces. This may suggest that these avatars may have triggered negative responses for female-identifying participants. This may be linked to existing research on avatar faces, features, and uncanniness which suggest that avatars that fail to display empathy or react appropriately may lead to assumptions of psychopathy in an avatar [11] and as such pose a threat. As such, one of the critical dimensions of uncanniness, discussed earlier, such as threat avoidance and alignment to terror management theory [17] as discussed may be more enhanced in female-identifying participants.

Further, it is also interesting to note that while despite some consistency in the avatars most uncanny ranks, the same consensus was not found the avatars ranked least uncanny. For both genders (sexes), two different female avatars were ranked as most uncanny.

These differences led to an additional investigation of the distribution of uncanniness score by the participants gender(sex).

While examining the distribution of scores, we see some interesting differences in the male and female participant scores for individual avatars. We see a bi-model distribution for the real humans and one of the Mid2-Low realism female avatars scores, which may suggest that the opinions of these avatars is divided. Of note are the scores for both the real humans who despite achieving high realism ranks, having their uncanniness scores fall roughly around the middle of the uncanniness rankings. When comparing the scores between the participants' gender(sex) these distributions were found to be statistically significantly different. Which may suggest that some participants could have been convinced that these avatars were computer-generated avatars rather than real humans. Further, we see that Bailie the one of the Mid2-Low realism female avatars was ranked as the least realistic but ranked only eighth on the uncanny ranks scores. This difference in ranking may indicate that the avatar was considered simultaneously unrealistic but not uncanny which is supported by existing literature lower realism levels which can lead to avatars being perceived as less uncanny [11].

The findings of this work highlight the importance of gender(sex) in avatar design decisions and how this variable may impact on the perceptions of avatar realism and uncanniness. As previously identified, a detailed investigation of the influence of gender(sex) in avatar realism and uncanniness perceptions is mostly missing in the decision making for avatar systems and from the current literature. Further, it is evident from the literature that the inevitable design choice of gender(sex) for an avatar has underlying cues and expectations placed on them based primarily on their perceived gender(sex). Together, the findings of the research provide key insights into gender(sex) based perception of avatars.

Although this work has made some significant contributions, it does have some limitations. First, the survey's sample has a limited ethnic diversity with the majority identifying as Asian or White/Caucasian backgrounds. This lack of diversity may lead to potential bias in the interpretation of the data. Second, the results may not be applicable to nonbinary genders(sexes) due to small sample sizes from these groups in this participant sample. However, this does present an avenue for future work. Third, there is some diversity in age (18–87 years old) however the mean age was ($M = 34.82$, $SD = 11.52$). The survey was unable to gather data from participants aged 18 and under as part of a restriction imposed by Mturk themselves. Thus, these findings may not be applicable to those who are under 18 years old.

Outside of the participant sample, the avatars also suffer from some limitations. Notably, the lack of ethnic diversity as they primarily have a homogenous Anglo-Saxon appearance with little distinction between their features. Lastly, the avatars are ranked by realism and uncanniness levels without the participants being given context for the avatars use. We note that perceptions might differ with context as previously identified [47]. Additionally, the complexity of the ranking task necessitated the use of a limited avatar set. However, given that avatars, as virtual human representations, could reflect the full diversity of the human form, it is arguable how expansive a set would be required to be representative, such a discussion is beyond the scope of this current work. However,

future work will seek to increase the avatar set evaluated to examine the differences in a more diverse set of avatars.

Our work has produced interesting insights into gender(sex) differences in the perception of avatars and generates many avenues for future research. First, we have focused on perceptual effects of the gender(sex) of both avatars and participants. Future analysis will also extend this to explore the differences in the rankings associated with avatar-participant self-similarity perceptions and avatar gender(sex). Additionally, another area for further analysis considers the individual attributes of each of the ten avatars through a gender(sex)-swapped lens. This will discuss how a simple gender(sex)-swap may function as a contributor to perceptions of avatar realism and uncanniness perceptions. In summary, the work presented here provides the basis for extending current knowledge of gender(sex) differences in the perceptions of avatar faces, regarding end-user perceptions of realism and uncanniness.

Acknowledgements. This research has been supported by an Australian Government Research Training Program (RTP) Scholarship. Additionally, the authors would like to thank Mr. Kim Colyvas from the University of Newcastle's Statistical Support Services, for his help and insights with the analysis of data for this project.

References

1. Gaba, D.M.: The future vision of simulation in health care. *Qual. Saf. Health Care* **13**(suppl 1), i2–i10 (2004)
2. Andrade, A.D., et al.: Avatar-mediated training in the delivery of bad news in a virtual world. *J. Palliat. Med.* **13**(12), 1415–1419 (2010)
3. Stumpf, S., et al.: Gender-inclusive HCI research and design: a conceptual review. *Found. Trends Hum Comput. Interact.* **13**(1), 1–69 (2020)
4. Wang, C.-C., Yeh, W.-J.: Avatars with sex appeal as pedagogical agents: attractiveness, trustworthiness, expertise, and gender differences. *J. Educ. Comput. Res.* **48**(4), 403–429 (2013)
5. Bailey, J., Blackmore, K., Robert, K.: Observing the Uncanny Valley: gender Differences in Perceptions of Avatar Uncanniness, In: 6th International Conference on Computer-Human Interaction Research and Applications. Valletta, Malta (2022)
6. Alexander, O., et al.: The digital emily project: achieving a photorealistic digital actor. *IEEE Comput. Graphics Appl.* **30**(4), 20–31 (2010)
7. Gleicher, M.: Animation from observation: motion capture and motion editing. *ACM SIGGRAPH Comput. Graph.* **33**(4), 51–54 (1999)
8. Nowak, K.L., Fox, J.: Avatars and computer-mediated communication: a review of the definitions, uses, and effects of digital representations. *Rev. Commun. Res.* **6**, 30–53 (2018)
9. Kang, S.-H., Watt, J.H.: The impact of avatar realism and anonymity on effective communication via mobile devices. *Comput. Hum. Behav.* **29**(3), 1169–1181 (2013)
10. Mori, M., MacDorman, K.F., Kageki, N.: The uncanny valley [from the field]. *IEEE Robot. Autom. Mag.* **19**(2), 98–100 (2012)
11. Tinwell, A.: *The uncanny valley in games and animation*. Boca Raton, AK Peters/CRC Press, FL, USA (2014)
12. Seyama, J.I., Nagayama, R.S.: The uncanny valley: Effect of realism on the impression of artificial human faces. *Presence Teleoperators Virtual Environ.* **16**(4), 337–351 (2007)

13. Zankaba, C., Goolkasian, P., Hodges, L.F.: Can a virtual cat persuade you? The role of gender and realism in speaker persuasiveness. In: Conference on Human Factors in Computing Systems. CHI 2006, pp. 22–27 April 2006. New York, NY, USA: ACM (2006)
14. Dobre, G.C., et al.: Nice is different than good: Longitudinal communicative effects of realistic and cartoon avatars in real mixed reality work meetings. In: CHI Conference on Human Factors in Computing Systems Extended Abstracts (2022)
15. Hudson, I., Hurter, J. Avatar Types Matter: Review of Avatar Literature for Performance Purposes. In: Lackey, S., Shumaker, R. (eds) Virtual, Augmented and Mixed Reality. VAMR 2016. Lecture Notes in Computer Science, vol 9740. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-39907-2_2
16. Tinwell, A., et al.: Facial expression of emotion and perception of the Uncanny Valley in virtual characters. *Comput. Hum. Behav.* **27**(2), 741–749 (2011)
17. MacDorman, K.F., et al.: Too real for comfort? Uncanny responses to computer generated faces. *Comput. Hum. Behav.* **25**(3), 695–710 (2009)
18. Tinwell, A.: 3D graffiti: Virtuality to the streets, In: Inopinatum: Urban Creativity, L.R. Borriello, C., Eds, Artigrafiche Bocchia: Rome. pp. 119–213 (2013)
19. Yoon, B., et al.: The effect of avatar appearance on social presence in an augmented reality remote collaboration. In: 2019IEEE Conference on Virtual Reality and 3D User Interfaces (VR). IEEE (2019)
20. Ambady, N., Bernieri, F.J., Richeson, J.A.: Toward a histology of social behavior: Judgmental accuracy from thin slices of the behavioral stream. In: Advances in experimental social psychology, pp. 201–271. Elsevier (2000)
21. Ambady, N., Gray, H.M.: On being sad and mistaken: mood effects on the accuracy of thin-slice judgments. *J. Pers. Soc. Psychol.* **83**(4), 947 (2002)
22. Schwind, V., Wolf, K., Henze, N.: Avoiding the uncanny valley in virtual character design. *Interactions* **25**(5), 45–49 (2018)
23. Davison, A.J., Deutscher, J., Reid, I.D.: Markerless motion capture of complex full-body movement for character animation. In: Computer Animation and Simulation 2001, pp. 3–14. Springer (2001)
24. Raj, A., et al.: ANR: Articulated Neural Rendering for Virtual Avatars. arXiv preprint [arXiv: 2012.12890](https://arxiv.org/abs/2012.12890), 2020
25. Seymour, M., et al.: Beyond deep fakes: Conceptual framework, applications, and research agenda for neural rendering of realistic digital faces. In: Proceedings of the 54th Hawaii International Conference on System Sciences (2021)
26. Zakharov, E., Ivakhnenko, A., Shysheya, A., Lempitsky, V.: Fast Bi-Layer Neural Synthesis of One-Shot Realistic Head Avatars. In: Vedaldi, A., Bischof, H., Brox, T., Frahm, J.M. (eds) Computer Vision – ECCV 2020. ECCV 2020. Lecture Notes in Computer Science(), vol 12357. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-58610-2_31
27. Fox and Bailenson: Virtual virgins and vamps: The effects of exposure to female characters’ sexualized appearance and gaze in an immersive virtual environment. *Sex Roles* **61**(3–4), 147–157 (2009)
28. Yee, N., Bailenson, J.: The Proteus effect: The effect of transformed self-representation on behavior. *Hum. Commun. Res.* **33**(3), 271–290 (2007)
29. Beltrán, M.: The importance of the avatar gender in training simulators based on virtual reality. In: Proceedings of the 18th ACM Conference on Innovation and Technology in Computer Science Education (2013)
30. Lehdonvirta, M., et al.: The stoic male: how avatar gender affects help-seeking behavior in an online game. *Games Culture* **7**(1), 29–47 (2012)
31. Hussain, Z., Griffiths, M.D.: Gender swapping and socializing in cyberspace: an exploratory study. *Cyberpsychol. Behav.* **11**(1), 47–53 (2008)

32. Baldwin, K.: Virtual avatars:trans experiences of ideal selves through gaming. *Markets Globalization Develop. Rev. (MGDR)* **3**(3) (2018)
33. GmbH., L., LimeSurvey: An Open Source survey tool. LimeSurvey GmbH, : Hamburg, Germany (2018)
34. amazon.com. amazon mechanical turk (2017)
35. Battocchi, A., Pianesi, F., Goren-Bar, D.: A first evaluation study of a database of kinetic facial expressions (dafex). In: *Proceedings of the 7th international conference on Multimodal interfaces*. ACM Trento, Italy, pp. 214–221 (2005)
36. Alexander, O., et al: *The Digital Emily Project: Achieving a Photoreal Digital Actor* (2017)
37. AppleInc., FaceShift. 2015, FaceShift AG
38. Metrics, I., Faceware. 2018, Faceware Tech
39. Nao4288, n., *Female Facial Animation ver 1.0 (No Sound)*. 2013
40. von der Pahlen, J., et al. Digital ira and beyond: creating real-time photoreal digital actors. In: *ACM SIGGRAPH 2014 Courses*. ACM (2014)
41. Lange, E.: The limits of realism: perceptions of virtual landscapes. *Landsc. Urban Plan.* **54**(1), 163–182 (2001)
42. Mantiuk, R.K., Tomaszewska, A., Mantiuk, R.: Comparison of four subjective methods for image quality assessment. *Comput. Graphics Forum* **31**(8), 2478–2491 (2012)
43. Winkler, S.: On the properties of subjective ratings in video quality experiments. in *Quality of Multimedia Experience*, In: 2009. QoMEx 2009. *International Workshop on Quality of Multimedia Experience*. USA. IEEE (2009)
44. Conover, W.J., Iman, R.L.: Rank transformations as a bridge between parametric and nonparametric statistics. *Am. Stat.* **35**(3), 124–129 (1981)
45. Brenton, H., Gillies, M., Ballin, D., Chatting, D.: The uncanny valley: does it exist. In: *Proceedings of Conference of Human Computer Interaction, Workshop on Human Animated Character Interaction*. Citeseer (2005)
46. Jentsch, E.: On the psychology of the uncanny (1906). *Angelaki: J. Theor. Humanit.* **2**(1), 7–16 (1997)
47. Rosen, K.R.: The history of medical simulation. *J. Crit. Care* **23**(2), 157–166 (2008)