View metadata, citation and similar papers at core.ac.uk
brought to you by CORE
provided by AIS Electronic Library (AISeL)

Proceedings of the 52nd Hawaii International Conference on System Sciences | 2019

Crossing the Uncanny Valley? Understanding Affinity, Trustworthiness, and Preference for More Realistic Virtual Humans in Immersive Environments

Mike Seymour
University of Sydney
mike.seymour@sydney.edu.au

Lingyao Yuan
Iowa State University
lyuan@iastate.edu

Alan R. Dennis
Indiana University
ardennis@indiana.edu

Kai Riemer
University of Sydney
kai.riemer@sydney.edu.au

Abstract

Developers have long strived to create virtual avatars that are more realistic because they are believed to be preferred over less realistic avatars; however, an "Uncanny Valley" exists in which avatars that are almost but not quite realistic trigger aversion. We used a field study to investigate whether users had different affinity, trustworthiness, and preferences for avatars with two levels of realism, one photo-realistic and one a cartoon caricature. We collected survey data and conducted one-on-one interviews with SIGGRAPH conference attendees who watched a live interview carried out utilizing two avatars, either on a large screen 2D video display or via 3D VR headsets. 18 sessions were conducted over four days, with the same person animating the photo realistic avatar but with different individuals animating the caricature avatars. Participants rated the photo-realistic avatar more trustworthy, had more affinity for it, and preferred it as a virtual agent. Participants who observed the interview through VR headsets had even stronger affinity for the photo-realistic avatar and stronger preferences for it as a virtual agent. Interviews further surprisingly suggested that our ability to cross the Uncanny Valley may depend on who controls the avatar, a human or a virtual agent.

1. Introduction

Virtual Reality (VR) is a form of visual and audio experiences that seek to immerse the user into a computer-mediated environment or a situation that simulates, yet is different from, the real world [5]. It is achieved by placing the viewer in a three-dimensional (3D) projected encapsulated space (typically via a headset), by using a stereoscopic two-dimensional (2D) display screen. A VR world can also be viewed on a regular monitor, but this reduces the interactivity. It can still, however, allow limited rotation of the camera view interactively. Headset viewed VR is more immersive than traditional human-computer interaction via a 2D screen, because the user is immersed in the projected reality and is free to move and explore the space from different viewing angles. This interaction between the viewer and the project reality is key, as it separates the immersive VR experiences from viewing on a computer screen where the viewer's position does not interactively affect the point of view of the scene [5]. VR can be free of digital characters, but much attention has been paid to improve the ability of the viewer to interact with virtual characters [5]. Such interactions range from the simplest form of observing the animated characters as a part of a predetermined scene to the most complex in which virtual characters, who are believably humans, interact with the viewer.

Voice-controlled digital assistants are currently popular in a wide range of consumer products, and nearly half of U.S. adults (46%) say they now use these applications to interact with smartphones and other devices [32]. Yet most of these devices present a disembodied voice as the representation of the assistant. Would interactions with these assistants change if they had a face and interacted like a human?

There has been a steady move towards creating characters and avatars that are more and more realistic [37, 38]. Much research has examined how users respond to more realistic characters or avatars [41]. The design of this study focuses on observing participants' interactions with human controlled avatars. Quantitative analysis was performed on human perceptions towards the avatars with different level of realism. However, we draw our discussions on human controlled agents versus Artificial Intelligence (AI)F agents from qualitative interviews with our participants.

The development of very realistic avatars is an important area of research but understanding how users react to these human-like digital entities is also important. Affinity and trust in online avatars and virtual agents are important factors that influence

URI: https://hdl.handle.net/10125/59615
ISBN: 978-0-9981331-2-6
(CC BY-NC-ND 4.0)

whether consumers visit and purchase from online retailers [8]. Do users have more affinity or trust for a virtual agent depicted using a highly realistic human avatar than one using a cartoon avatar, or would they prefer one agent over the other? Understanding these issues have both theoretical and practical implications, as developers spend millions to push such technologies forward, as companies make deployment decisions, and as users begin to encounter such avatars. This study strives to address two questions:

RQ1: Are there differences in user perceptions of (i.e., trustworthiness, affinity) and preferences for avatars with different levels of realism?

RQ2: Does the virtual environment (i.e., immersive 3D or traditional 2D) impact user perceptions of, and preferences for, avatars of different levels of realism?

2. Theoretical Background

VR has moved from research curiosity and gaming platform to "gain legitimacy in business and educational settings for their application in globally distributed, project management, online learning and real-time simulation" [36]. Until recently there was little significant organizational application [35], so very little of VR research, has "found its way into IS research"[36]. This has changed with the introduction of inexpensive consumer-grade VR headsets that has generated new interest in enhancing existing systems and create new opportunities.

Research suggests that users may see the avatar either as a direct extension of the user or as something separate and distinct [35]. At the heart of the experience is the issue of agency and whose identity the observers believed they are experiencing. While the avatars are a mix of realism of their driving participants, they also exist simultaneously as fantastical representations, being able to look and act differently than the person controlling by them.

2.1. The Uncanny Valley and Affinity

The 40-year-old Uncanny Valley theory [31] plays a key role in research on users' reactions to avatars and agents. The theory argues that users have greater affinity for avatars that are more realistic. User affinity increases as the avatar becomes increasingly realistic, until the avatar is semi-realistic, at which point affinity drops dramatically because a partially realistic avatar triggers unease in users. See Figure 1. As realism increases, there comes a point where the valley has been crossed and the avatar's affinity increases to its highest level. It does not require the realistic avatar to be imperceptibly real, just very close. Thus, "crossing the Uncanny Valley" has attracted much research and commercial attention.

The Uncanny Valley uses the concept of "affinity", which comes from an original Japanese word, Shinwakan (親和感), and thus is open to interpretation as it is translated into English. "Affinity" has emerged as the preferred translation [31, 41]. Affinity is an indicator of whether an avatar is in or across the Uncanny Valley. The theory is not based on empirical data, just conjecture. It also predicts a magnified effect when viewing the target with movement over a still image.

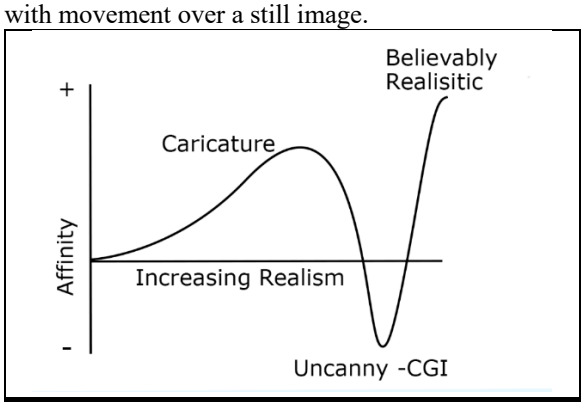

Figure 1. Uncanny valley

The cause(s) of the Uncanny Valley are not clear, but there are many different theories (see [41] for a summary). Three theories are particularly relevant for our research. The first theory argues that the drop in affinity in the Uncanny Valley is due to perceptual surprise [29]. In the first 100-300ms after seeing what could be a face, our subconscious initially concludes that the almost-human avatar is a human and creates an expectation of its humanity. It then directs our conscious attention to focus on it. Our conscious attention is surprised when it determines that the avatar is actually not a human and this surprise triggers a negative emotion.

A second theory argues that we perceive the almost-human avatar to be human, but its less than perfect features lead us to dehumanize it [41]. Dehumanization is the process whereby we perceive a human to lack the attributes that comprise what it means to be a human. It occurs when we see a person as a member of an out-group that is different from the in-group of people like ourselves; they become animals (less intelligent) or machines (lacking emotions) [14]. In either case, this dehumanization triggers negative emotions.

A third theory is based on evolution and argues that our responses to almost-human avatars are subconscious reactions for self-preservation [31]. We perceive almost-human avatars to be humans

exhibiting a psychopathic personality disorder [39]. These almost-human avatars are perceived to be callous and dishonest because they fail to accurately display emotions and/or behave in the same way as healthy humans.

A key point in all these theories is that they argue that affinity for the avatar is not deliberate; the shared conclusion is that affinity is driven by subconscious processes that are beyond conscious control. The first two theories are based on visual perceptions triggering subconscious processes, so a static image is sufficient to trigger our aversion. The third theory argues that behavior that triggers aversion, so the avatar must be interacting; a static image is not sufficient.

Empirical studies that have examined the Uncanny Valley primarily have used static images or scripted video clips; few have explicitly explored interactivity [37], so, we have little understanding of how users perceive interacting avatars. The human face plays an important role in communication [37]; much information is communicated nonverbally by our facial expressions [42]. Cartoons lack detailed facial muscles, so they have a much narrower array of nonverbal signals they can communication. We theorize that more human-realistic avatars have the potential to improve communication with virtual agents. After all, the Uncanny Valley theory argues that close to human-realistic avatars should engender more affinity than cartoon avatars [31], but we need to cross the Valley. This leads to our first proposition:

> ***Proposition 1.*** *Users will have greater affinity for a human-realistic avatar than a cartoon avatar.*

2.2. Trustworthiness

Trust is an individual's willingness to be vulnerable to the actions of the other for a particular action, irrespective of the trustor's ability to monitor or control the trustee [4, 25]. Trustworthiness is an assessment of whether another person or thing is worthy of trust [25]. Trust is between people [25], but also applies to information systems [3, 20, 23, 46].

Mayer, et al. [25] argue that trust is a function of the trustor's disposition to trust and the trustor's assessment of the trustee's ability, integrity, and benevolence. Trust is refined through interaction [21, 25]. The trustor's disposition to trust is independent of the trustee; it is a "generalized attitude" learned from experiences of fulfilled and unfulfilled promises [31, 34], and varies from person to person.

The other three elements of trust are based on the trustor's assessment of the trustee [16, 25, 33]. Ability refers to the skills that enable the trustee to be competent within some specific domain. Ability is key, because the trustor needs to know that the trustee is capable of performing the task he or she is being trusted to do. Integrity is the adherence to a set of principles that the trustor finds acceptable. Integrity is important because it indicates the extent to which the trustee's actions are likely to follow the trustee's espoused intentions. Benevolence is the extent to which the trustee is believed to feel interpersonal care, and the willingness to do good, aside from a profit motive. Benevolence is important over the long term, because it suggests that the trustee has some attachment to the trustor, over and above the transaction in which trust is being conferred.

Ability and integrity may be more important than benevolence when the task is transaction-oriented because the trustor just needs to have confidence that the trustee has the ability to complete the transaction [11]. For advice giving or recommendations, benevolence may be more important because to provide good advice and recommendations the trustee must take into account the trustor's best interests, separate from a profit motive.

Benevolence and integrity are human characteristics [11]. While we can think of machines as having an ability to perform a task, they lack the fundamental capability to adhere to principles (integrity) or feel interpersonal care (benevolence). Therefore, we theorize that human-realistic avatars are more likely to be perceived as having integrity and benevolence than cartoon avatars that are clearly non-human. Because integrity and benevolence affect trustworthiness, we theorize that human-realistic avatars will be perceived as more trustworthy than cartoon avatars. We also theorize that this will hold between human-realistic avatars and lesser realistic human avatars that lie in the Uncanny Valley. Thus:

> ***Proposition 2.*** *Users will ascribe greater trustworthiness to a human-realistic avatar than to a cartoon avatar.*

2.3. User Preferences

Affinity and trustworthiness are two important characteristics of virtual agents [8]. Affinity has often been linked to increased preferences for interaction with avatars and web sites in general [6, 8, 22]. Likewise, trustworthiness is an important factor influencing both interpersonal preferences and preferences for websites – and increased interactions with both [11, 26]. We argued above that human-realistic avatars would induce greater affinity (Proposition 1) and greater trustworthiness (Proposition 2) than a cartoon avatar. Taken together, we theorize that human-realistic avatars should be

preferred as virtual agents over cartoon avatars. Thus:

Proposition 3. *Users will prefer a human-realistic avatar to a cartoon avatar as virtual agent.*

2.4. Display Format

There are two fundamentally different ways in which VR can be used. One is an immersive 3D environment, which is typically provided by using a 3D VR headset. The second is by projecting the virtual world onto a flat 2D display screen. The 3D headset differs in two theoretically different ways from the 2D screen. First, the 3D headset enables the user's view of the world to change as the user moves his/her head or moves around physically. The user is able to peer around objects to see them from a different vantage point, in the same way that moving in the physical world changes the user's view. Second, the 3D headset ensures the users only see the virtual world. Unlike the 2D screen which enables users to see other objects in their physical world (e.g., their desk), the 3D headset masks the user's physical world so that he or she can only see the virtual world. We theorize that these two theoretical mechanisms will strengthen the effects of virtual experience. This will heighten the differences between the realistic avatar and the cartoon avatar.

Previous research comparing 2D VR presentation on screens with 3D VR headsets have shown some important differences. Ashraf et al. [2] briefly summarize prior research and report on a randomized experiment comparing laparoscopic surgery using 2D screens and 3D headsets. This study, along with prior research on the use of VR in surgery and surgical training, suggests there may be some improvement in skills (e.g., faster times and fewer errors) when using 3D headsets. It is important to note that these tasks require direct physical interaction in a three-dimensional environment, which our study does not. Nonetheless, we propose:

Proposition 4. *Individuals who view avatars using immersive 3D virtual reality headsets a) will feel more affinity towards the human-realistic avatar, b) will ascribe greater trustworthiness to a human-realistic avatar, and c) will be more likely to prefer a human-realistic avatar.*

3. Research Methodology

We conducted a field study at the SIGGRAPH Conference 2017 held in Los Angeles from July 30th to August 3rd. The event was an invited and curated part of the Conference and constructed with the resources of a range of industry and academic partners. We conducted 18 sessions over four days, collecting quantitative surveys from and doing qualitative interviews with audience members. We first describe the event and then discuss the data collection.

3.1. MEETMIKE Event Description

MEETMIKE featured Mike Seymour interviewing 18 leading experts in the field of digital human technology in real-time utilizing a human-realistic avatar ("Digital MIKE") in a "virtual studio setup in Sydney". The event was presented as part of the conference's VR Village, (see Figure2).

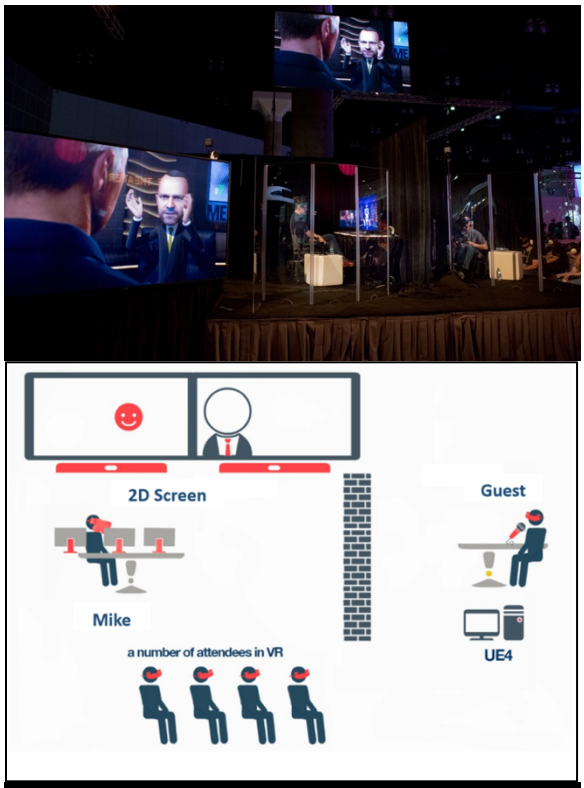

Figure 2. MEETMIKE Event Image (Above) and Design (Below)
Photograph Courtesy of the Authors

There were four roles:

1) The Host, Mike Seymour, was conducting interviews. Digital Mike, a highly realistic virtual avatar, was developed based on Mike Seymour;

2) The Guest in each session was a well-known industry expert working in visual design and/or the movie industry. Each guest participated only once, so there were 18 different Guests, one for each session. Each guest was represented by a unique cartoon avatar that was custom-designed to be a caricature of the

guest, so there were 18 different cartoon avatars.

3) The VR audience members were four SIGGRAPH conference participants who were pseudo randomly chosen to observe the interview in VR using VIVE headsets;

4) The non-VR audience members were SIGGRAPH conference participants who observed the interview via traditional 2D monitors. Audience size varied but was usually about 30 people.

Each of the 18 sessions lasted about 20 minutes. The Host and the Guest had the active roles carrying on a conversation on the history, progress and the future of virtual human technology. The VR audience and non-VR audience were observers of this conversation. The event environment was a constructed space at the Conference that allowed two participants, the Host and Guest, to sit on either side of a barrier and only see and hear each other via the VR technology.

Figure 3. Head rig with stereo cameras (HMC)(left) and computer generated eye close up (right).
Photograph Courtesy of the Authors

The Host was presented as a human-realistic avatar. The Host wore a Head Mounted Camera Rig (HMC) with two stereo computer vision cameras which enabled stereo 3D reconstruction of the Host's expressions and the 'solving' of the Host's expressions into 'expression space' (Figure 3). The expression space is based on the Facial Action Coding System (FACS) system of expressions. This allowed subtle expressions on the Host's face to be interpreted into a set of computer instructions that drove a fully 3D computer generated avatar of the Host in real time. This avatar model was displayed only from the chest up. The system mapped head movement and detailed facial expressions from the host to the digital avatar. This digital avatar was built based on extensive scanning of Mike Seymour's face and research that is outside the scope of this paper [37]. Creation of the avatar involved extensive and custom state of the art Game Engine tools (developed in Epic Games' UE4) to produce a professional digital avatar with precise features and real-time facial responses. The motion of the avatar was driven by a pair of stereo computer vision cameras worn by the Host, augmented by a VIVE capture volume for head movement (using a VIVE 'puck' mounted on the HMC). Complex custom specialist code, deep learning face tracking techniques were used on the Host to produce the highest fidelity possible facial input data from Industry partner Cubic Motion. This input was then interpreted into the rendered expressions of Digital Mike. Digital Mike's face had an extensive range of emotion and state of the art expression realism due to a custom facial rig developed by 3lateral. Due to the complexity of the model, and quality of the textures and rendering, even with the most complex hardware at our disposal, only a chest up character could be rendered at the desired 90 frames per second required VR rate.

The Guest was presented as a cartoon avatar. The Guest's avatar was based on a single jpeg image of the guest provided in advance, using custom AI technology from industry partner Loom.ai. The Guest wore a VR headset, which had been specially modified to provide stereo eye and mouth tracking, via the addition of two sets of small stereo computer vision cameras. This headset enabled the Guest to experience the experiment in VR, but at the cost of a lower complexity and fidelity avatar. The Guest avatar provided tracked head and arm movements that enabled the Guest to speak, move, and produce hand gestures. The Guest's cartoon avatar used estimated facial expressions, created from each Guest using deep learning extrapolation from the reduced input of just mouth and eye positions. For all VR participants and the Host, these inputs were done in VIVE capture volumes that enabled the characters to be rendered in the virtual space with the correct head movement in real-time. The small audio delay due to processing was adjusted to maintain lip sync.

Figure 2 shows an example of the Host's avatar and a Guest avatar as seen in VR and the real Host and a real Guest. Audience members wearing VR headsets could only view the avatars in VR. Audience members watching on the 2D display saw the same VR interaction, but these audience members could shift their gaze between the 2D display and the real host and guest who were visible on stage.

The interactions between the Host and the Guest were rendered in real time at 90 fps in VR and at 2K resolution (Figure 4). For either of the Host and the Guest, two computers (so four computers in total) were dedicated to providing real time interactive facial and eye tracking with high resolution visualizations.

Nine high-end PC computers (8 and 10 core, 32Gig RAM PCs) with 1080 NVIDIA GPU graphics cards were divided up: two PCs for either of the Guest and Host, one for each 4 VR audience (allow them to customize their view or perspective), and one for the general audience to watch (at a different 60fps and quality settings) (Figure 5).

Figure 4. Host avatar (upper left) ,sample guest avatar (upper right) with real host (lower left) and real guest (lower right)
Photograph Courtesy of the Authors

3.2. Data Collection

3.2.1. Surveys. Surveys were distributed at the end of each of the 18 sessions. 157 valid surveys were collected. 43% were VR audience wearing the VR headset. 68% of the respondents were male and 71% were Caucasian. Scales for affinity and trustworthiness were adopted from prior research and modified for this study. Cronbach's alpha of trust items towards Digital Mike is 0.92 and towards the Guest avatar is 0.95. Cronbach's alpha of affinity items towards Digital Mike is 0.82 and towards the Guest avatar is 0.80.

Participants were asked to choose between the Guest avatar and Digital Mike as their preference for a virtual agent using one 7-item question on the survey. "Suppose you were to use a virtual concierge. Which type of concierge would you prefer: the caricature used by the Guest or the realistic avatar used by Mike?" The scale went from Guest on the left to Mike on the right, with the midpoint as Neutral.

Figure 5. VR Audience Image
Photograph Courtesy of the Authors

Participants were also asked about the familiarity with MIKE and with the Guest on a 4-point scale (included as the control variable). Demographic information, including gender and ethnicity, was also collected, because some individuals display face-blindness for individuals of other races [40].

3.2.2. Interviews. Thirty-two one-on-one qualitative interviews were conducted with a goal of understanding participants' perceptions of the two avatars and an imagined, soon to be enabled, reality where these avatars could represent virtual agents. The qualitative interviews were conducted immediately after their experience and lasted approximately five and half minutes on average. Twenty participants (62.5%) viewed the event on the 2D screens, ten (31%) used the VR headsets, and the remaining two participants (6%) were Guests. Two thirds were male (66%), which reflects the fact the conference is predominantly attended by males. The average age of participants was approximately 36. All were adults of a working age (over 20 and less than 60). The interviewer was an experienced qualitative academic researcher, and the interviews took place normally within minutes of the session finishing. The participants were asked similar questions ranging from general questions, such as asking the participants to describe what they had just witnessed, to more specific questions, such as their view on the usefulness or applicability of this technology in their work context.

4. Results

4.1. Quantitative Surveys

4.1.1. Analysis Technique. We used standard General Linear Methods (GLM) to analyze the preference for avatars data. We used Hierarchical Linear Model

(HLM) [15] to analyze the data on trustworthiness, and affinity. HLM is a form of regression that considers multiple levels of analysis in one statistical equation, where traditional regression techniques are not appropriate due to nested data [1, 19]. The lowest level (level 1) of the HLM model is the avatars with different level of realism; the second level (level 2) is participant level characteristics including whether the participant wore a VR headset or not. The third level (level 3) is the session level, controlling for underlying characteristics of the Guest that could impact the constructs of interest.

4.1.2. Affinity. Table 1 presents the results. The intercept term on the Avatar is significant (p=.000) and positive, meaning participants had more affinity toward Mike than the Guest, supporting Proposition 1. VR is significant (p=.015) and positive, meaning people wearing VR headsets rated Mike with more affinity than the Guest, supporting Proposition 4a.

The other terms in the Avatar equation are not significant, meaning that affinity for one avatar or another is not affected by familiarity with Mike or the Guest, gender or ethnicity. Several terms in the Intercept equation are significant, which mean they have main effects. Different sessions resulted in different affinity for both avatars and regardless of whether the participant was in VR environment or not. Participants' familiarity with the Guest is significant (p=.008) and positive, indicating people who were familiar with the guest rated both Mike and the Guest as having higher affinity than people who didn't know the Guest. Gender approached significance (p=.056) and is negative, meaning that males may or may not have rated both Mike and the Guest as having lower affinity. Ethnicity is significant (p=.016) and negative, meaning white people (i.e., people of the same race as Mike) rated both Mike and the Guest as having less affinity than people of non-white races.

4.1.3. Trustworthiness. Table 1 also presents the results for trustworthiness. The intercept on the avatar is significant (p=.004) and positive, meaning participants rated Mike as more trustworthy than the Guest, supporting Proposition 2. VR is not significant (p=.902), meaning wearing VR headsets did not affect trust, counter to Proposition 4b. With one exception, the other terms in the Avatar equation are not significant, meaning that trustworthiness is not affected by familiarity with Mike, gender or ethnicity. Familiarity with the Guest was significant (p=.017) and negative, indicating that those with greater familiarity with the Guest had less trust in Mike, but this is offset by a significant (p=.025) positive main effect for familiarity with the Guest meaning participants who were familiar with the Guest rated both Mike and the Guest as being more trustworthy than people who didn't know the Guest; taken together, these two terms show that participants who were familiar with the Guest, rated the Guest as having higher trustworthiness but not Mike (-.191 and .259, combined effect for Mike =.068).

Table 1. HLM Results

	Affinity		**Trustworthiness**	
Level 1 Level 2	Coeffi cient	*p* value	Coeffi cient	*p* value
Intercept				
Intercept	4.686	0.000	5.001	0.000
VR	0.068	0.725	0.297	0.155
Familiarity-Mike	0.086	0.269	-0.072	0.475
Familiarity-Guest	0.245	0.008	0.259	0.025
Gender	-0.294	0.056	0.097	0.481
Ethnicity	-0.433	0.016	-0.465	0.003
Avatar				
Intercept	0.898	0.000	0.437	0.004
VR	0.525	0.015	0.028	0.902
Familiarity-Mike	-0.063	0.633	0.079	0.483
Familiarity-Guest	-0.156	0.172	-0.191	0.017
Gender	0.152	0.587	0.147	0.460
Ethnicity	0.069	0.804	0.025	0.893

4.1.4. Preference as Virtual Agent. We used GLM to analyze the preference as virtual agent results. A -3 indicated the participant strongly preferred the Guest avatar and a +3 strongly preferring the Mike avatar. The overall mean was 1.45, which was significantly greater than zero (p=.000), thus providing support for Proposition 3. We split the data into two groups, those wearing VR headsets and those viewing on the 2D screen. Results show that both groups significantly preferred the Mike avatar to the Guest avatar for a virtual agent (VR=1.81, 2D=1.19; p=.000). There were significant differences between the two groups (p=.046), supporting Proposition 4c.

4.2. Qualitative Interviews

The aim of qualitative research was to provide a richer understanding of the same issues in the quantitative research. Interviews were done at the same time and under similar conditions as the surveys. The interviews were recorded and transcribed. They were then examined in NVivo (v11) for both broad thematic issues and any unanticipated responses.

4.2.1. Affinity, Trustworthiness, and Preferences. The qualitative data reinforced the quantitative data. Interviewees reported more affinity for the photorealistic avatar than the cartoon avatar and saw it as more trustworthy. More interviewees preferred the photorealistic avatar to the cartoon avatar. These results are useful as they provide a different viewpoint that triangulates well with the quantitative data. However, there were two additional insights.

4.2.2. Avatars versus Humans. Respondents shifted between seeing the session as interactions between avatars and interactions between the humans controlling the avatars. In some cases, the avatars were spoken of as separate from the humans, while in others, the avatars stood in place of the real humans. Both Mike and the Guest's prior reputations and activities enabled some respondents to have some level of familiarity with one or both of them. When discussing appearance, respondents saw the avatars as extensions of the humans (e.g., how "real" the Mike avatar looked). However, in speaking of the topic discussions and emotional responses to the experience, the respondents' language shifted to seeing the avatars as separate from their human controllers. In appearance, the avatar was seen as a technical reflection of the human controlling it, while in emotional response, the avatars were the source of the emotion, not the humans. When asked to comment on the technology, respondents saw the avatar as a stand-in for the human, but when asked to comment about the interaction (absent a reminder about technology), respondents saw the avatar as the actor and its human controller disappeared into the background.

This situation may be a good embodiment of Goffman's [13, 12] dramaturgical framing of social interaction as theater. Although based on face-to-face communication among humans, Goffman's work provides a useful vocabulary for describing interaction among avatars, particularly the portion that segments interaction into "front stage" and "backstage." Front stage behavior is characterized by the presence of an "audience," individuals who expect one's actions to be consistent with an official role in its relationship to the audience. Backstage behavior is characterized by interactions among "teammates," people who share the same role with respect to the audience.

In our study, the avatars were the front stage and the humans were the backstage controllers. For those viewing on the 2D screens, the front and back stages were simultaneously present, the front stage on the 2D video display and the backstage actually physically present in their visual field. For those viewing on the VR headsets, only the front stage was visually present.

Our respondents recognized the distinction between the front stage avatars and the backstage humans controlling them. Yet the distinctions were the strongest when discussing the technology, which forced them to separate front stage from back stage. The distinctions blurred or disappeared when they discussed the interaction – respondents appeared to focus on the front stage and overlook the backstage.

4.2.3. The Uncanny Valley from a Dramaturgical Frame. Previous theories to explain the Uncanny Valley effects are grounded in the issue of image fidelity [41]. The essence of these theories is that the avatar is an imperfect rendering of a human, and thus our subconscious triggers an aversive reaction because it perceives the avatar as a psychopath [39], it is surprised [29], or it dehumanizes the avatar [41].

Goffman's [13, 12] dramaturgical framing helps us understand what was obvious – at times – to our respondents: the avatars on the front stage were separate from the humans controlling them from the backstage. But what happens when we are unsure about what is controlling the avatar? Is it an avatar being controlled by a human or is it a non-human virtual agent controlled by artificial intelligence (AI)?

Our interviews suggest there may be emotional bias against dealing with a realistic-looking avatar that is an artificial virtual agent controlled by AI. It is this awareness of the lifelike yet artificial presence that several respondents expressed concerns about and wanted to avoid. A typical comment, from those who expressed reservations when invited to extrapolate on the future AI uses, was that the realistic human looking MIKE avatar if not driven by an actual human, would "creep you out, but at the same time it is really cool."

Some went further. When asked to imagine the technology as the user interface of a virtual assistant such as Apple's Siri, one interviewee replied, "I don't think I would want to see a super real face. I feel like I would be more comfortable with a distinction between me and her". Another said that a realistic face would be something they would like to see on an assistant, but it would be "a little confusing", due to the lack of clarity between what was human and AI.

Finally, a couple of interviewees rejected the notion. While they responded positively to the avatar driven by a real person, they speculated that they if this had been driven by AI, they would "possibly find it creepy" and it would "probably be too much". One commented that it would be a "bit spooky". This sentiment was a minority opinion, but, it is important to note that the sample was drawn from SIGGRAPH attended by people who are technically literate and positively inclined to new technology.

This discomfort arose only when the realism of

the avatar approached a near perfect human form. There were no concerns about the avatars displayed using cartoon caricatures. We conclude that at lower levels of realism, this lack of perfect reproduction avoided a sense of deception and thus there were no issues with affinity. However, once the avatar becomes highly realistic, users may find the lack of knowing who or what is backstage controlling it as unsettling as prior Uncanny Valley visual responses. We speculate that a highly realistic human looking avatar controlled by AI would generate a sense of unease because your subconscious would perceive the avatar as human, but your conscious would know it was not, thus creating cognitive dissonance.

This theoretical framing leads to very different predictions for our ability to cross the Uncanny Valley. As with past theories, this framing would lead us to conclude that affinity would increase as the realism of front stage avatars increases until we reach the Uncanny Valley. The ability to cross the Uncanny Valley depends on the backstage controller. If the backstage controller is human, then increasing realism will enable us to cross the Valley. If the backstage controller is AI, then we may never cross the Valley for some users; increasing the realism of the interactive character will only increase our cognitive dissonance leading to lower affinity.

5. Discussion

In summary, our results show that participants had greater affinity for the more human-realistic avatar than the cartoon avatar, perceived the human-realistic avatar to be more trustworthy, and preferred it as a virtual agent. Participants wearing VR headsets (as contrasted with those watching a 2D display) had even stronger affinity for the more human-realistic avatar and were more likely to prefer it as a virtual agent. These results would suggest that in this case, the more human-realistic avatar successfully crossed the Uncanny Valley, although our interview results suggest some cautionary caveats to this conclusion.

Humans are hard wired to interpret human faces. Our brains can read faces with far more fidelity than any other object. Evolution has left us with the ability to quickly identify and reject artificial faces which are only approximately close to realistic [27]. Not only can we detect these inferior renditions but we unconsciously react to them far less favorably than a simple caricature [31]. As VR and Augmented Reality (AR) become more common, it will become important to ensure that the human faces we see in these environments do not trigger aversion associated with the Uncanny Valley. We believe that our research indicates that we are on the cusp of crossing the Uncanny Valley, although it also suggests some important limitations.

The more realistic avatar was perceived to be more trustworthy than the cartoon avatar. Trustworthiness is an important factor in both interpersonal interaction [25] and interaction with technology artifacts [20, 23]. Our avatars were technology artifacts controlled by humans and designed to induce a perception of humanness, so trustworthiness is important, regardless of whether they are perceived to be technology, human, or a bit of both. We argued that one fundamental theoretical difference was the potential for the more realistic avatar to be perceived to have more integrity and more benevolence than an artificial cartoon which in turn would increase the perceptions of trustworthiness. Our results provide some support for these arguments.

Our participants could distinguish between the front stage avatar and the backstage controlling human, but this distinction blurred as discussion moved from the technology to the emotional effects (e.g., affinity). Survey participants reported they would prefer the more realistic avatar as a virtual agent, but those interviewed raised concerns about a realistic-looking virtual agent controlled by AI that was not human. We conclude that we can cross the Uncanny Valley when avatars are controlled by humans. However, our interviews offer a new theoretical argument that challenges whether virtual agents (i.e., non-human avatars) can ever completely cross the Uncanny Valley for some people.

Interestingly, whether participants viewed the interaction using VR headsets or on a 2D screen affected affinity and preferences, but not trustworthiness. The VR headsets obscured the backstage, while the front stage and backstage were always visually present when using the 2D screens. We speculate that affinity and preference may be more surface emotions than trustworthiness which requires more thought; thus, they may be more strongly influenced by the viewing media.

One major limitation is that is an initial field study, rather than a controlled laboratory study. We did not vary the avatar of the Host because it was technically difficult to create even one highly realistic avatar. Thus, we could not randomly assign the human controller to the avatar as in a controlled experiment. We attempted to mitigate this issue by using 18 different Guests, each with their own cartoon caricature. However, the effects we observed could simply be due to underlying differences in affinity, trustworthiness, and preferences for the human controllers (i.e., Mike and the 18 individual Guests), not the avatars; we controlled for familiarity with both Mike and the Guest. Mike Seymour, the Host

participant, was not more qualified or more well-known than the expert Guests. Nonetheless, more research in controlled laboratory settings is needed. The second major limitation is that the participants for this study were attendees at the leading graphics conference. We selected these participants because they are familiar with VR and thus are not likely to experience a novelty effect as might the general population. We wanted to research the digital humans not research the broader experience of seeing cutting edge graphics in VR. We also need the audience to have a similar perspective on the discussed topics. A completely random community could include people with no interest in the topic and thus their general disinterest might cloud their answers on trust.

Despite these limitations, our qualitative results suggest an alternative theory for the Uncanny Valley and raise some serious limitations on our ability to cross it. One important step for future research would be explore the role of the backstage actors in influencing the Uncanny Valley. Our participants knew the front stage avatars were controlled by backstage humans and were not AI controlled virtual agents. If the participants believed that the front stage avatars were controlled by backstage AI, could these avatars cross the Uncanny Valley? We need more research to test this theoretical proposal that it is not only what is visible on the front stage, but also the backstage controller, that will influence affinity and our ability to cross – or not cross – the Valley.

Our results also suggest that VR headsets matter. We need more research to better understand why. Is it because VR headsets make the environment more immersive or seem more real? Or is it because in our study VR headsets removed the backstage from view, and thus strengthened the perceptions of the avatar as an entity separate and distinct from its controller? If so, then a 2D screen that also removed the backstage from view would have similar effects.

What does this mean for VR developers and for companies looking to deploy VR and virtual agents? First, users have more affinity for and trust in photo realistic avatars than cartoon avatars and prefer them to cartoon avatars. Thus, we recommend that developers implement more photo realistic avatars. This may be tempered to some extent by the application. Our research examined avatars controlled by humans (e.g., for social or gaming). Our surveys showed that our participants preferred photo realistic avatars as virtual agents, although interviews with our participants suggest that these effects may or may not generalize to agents controlled by AI (e.g., cognitive agents). Second, the way in which users view the avatar is important; we recommend that organizations consider VR headsets for such applications.

6. References

[1] S. Ang, S. Slaughter and K. Yee Ng, "Human capital and institutional determinants of information technology compensation: Modeling multilevel and cross-level interactions", Management Science, 48 (2002), pp. 1427-1445.

[2] A. Ashraf, D. Collins, M. Whelan, R. O'Sullivan and P. Balfe, "Three-dimensional (3D) simulation versus two-dimensional (2D) enhances surgical skills acquisition in standardised laparoscopic tasks: a before and after study", International Journal of Surgery, 14 (2015), pp. 12-16.

[3] I. Benbasat and W. Wang, "Trust in and adoption of online recommendation agents", Journal of the association for information systems, 6 (2005), pp. 4.

[4] G. A. Bigley and J. L. Pearce, "Straining for shared meaning in organization science: Problems of trust and distrust", Academy of management review, 23 (1998), pp. 405-421.

[5] G. C. Burdea and P. Coiffet, *Virtual reality technology*, John Wiley & Sons, 2003.

[6] A. Cafaro, H. H. Vilhjálmsson and T. Bickmore, "First Impressions in Human--Agent Virtual Encounters", ACM Transactions on Computer-Human Interaction (TOCHI), 23 (2016), pp. 24.

[7] D. DeVault, J. Mell and J. Gratch, *Toward natural turn-taking in a virtual human negotiation agent*, *AAAI Spring Symposium on Turn-taking and Coordination in Human-Machine Interaction. AAAI Press, Stanford, CA*, 2015.

[8] R. Etemad-Sajadi, "The impact of online real-time interactivity on patronage intention: The use of avatars", Computers in Human Behavior, 61 (2016), pp. 227-232.

[9] S. Finkelstein, E. Yarzebinski, C. Vaughn, A. Ogan and J. Cassell, *The effects of culturally congruent educational technologies on student achievement*, *International Conference on Artificial Intelligence in Education*, Springer, 2013, pp. 493-502.

[10] K. Franceschi, R. M. Lee, S. H. Zanakis and D. Hinds, "Engaging group e-learning in virtual worlds", Journal of Management Information Systems, 26 (2009), pp. 73-100.

[11] D. Gefen and D. W. Straub, "Consumer trust in B2C e-Commerce and the importance of social presence: experiments in e-Products and e-Services", Omega, 32 (2004), pp. 407-424.

[12] E. Goffman, *Frame analysis: An essay on the organization of experience*, Harvard University Press, 1974.

[13] E. Goffman, *The Presentation Of Self In Everyday Life." Garden City, NY: Doubleday And Company*, Inc, 1959.

[14] N. Haslam and S. Loughnan, "Dehumanization and infrahumanization", Annual review of psychology, 65 (2014), pp. 399-423.

[15] D. A. Hofmann, "An overview of the logic and rationale of hierarchical linear models", Journal of

management, 23 (1997), pp. 723-744.
[16] S. L. Jarvenpaa and D. E. Leidner, "Communication and trust in global virtual teams", Journal of Computer-Mediated Communication, 3 (1998).
[17] F. Kaba, "Hyper realistic characters and the existence of the uncanny valley in animation films", International Review of Social Sciences and Humanities, 4 (2013), pp. 188-195.
[18] O. Klehm, F. Rousselle, M. Papas, D. Bradley, C. Hery, B. Bickel, W. Jarosz and T. Beeler, *Recent advances in facial appearance capture*, *Computer Graphics Forum*, Wiley Online Library, 2015, pp. 709-733.
[19] D.-G. Ko and A. R. Dennis, "Profiting from knowledge management: the impact of time and experience", Information Systems Research, 22 (2011), pp. 134-152.
[20] S. Y. Komiak and I. Benbasat, "The effects of personalization and familiarity on trust and adoption of recommendation agents", MIS quarterly (2006), pp. 941-960.
[21] R. Lewicki and B. Bunker, *Conflict, cooperation and justice, chapter Trust in relationships: a model of trust development and decline*, Jossey-Bass, 1995.
[22] E. T. Loiacono, R. T. Watson and D. L. Goodhue, "WebQual: An instrument for consumer evaluation of web sites", International Journal of Electronic Commerce, 11 (2007), pp. 51-87.
[23] P. B. Lowry, A. Vance, G. Moody, B. Beckman and A. Read, "Explaining and predicting the impact of branding alliances and web site quality on initial consumer trust of e-commerce web sites", Journal of Management Information Systems, 24 (2008), pp. 199-224.
[24] M. B. Mathur and D. B. Reichling, "Navigating a social world with robot partners: A quantitative cartography of the Uncanny Valley", Cognition, 146 (2016), pp. 22-32.
[25] R. C. Mayer, J. H. Davis and F. D. Schoorman, "An integrative model of organizational trust", Academy of management review, 20 (1995), pp. 709-734.
[26] D. H. McKnight, L. L. Cummings and N. L. Chervany, "Initial trust formation in new organizational relationships", Academy of Management review, 23 (1998), pp. 473-490.
[27] M. Meng, T. Cherian, G. Singal and P. Sinha, "Lateralization of face processing in the human brain", Proceedings of the Royal Society of London B: Biological Sciences (2012), pp. rspb20111784.
[28] T. Merel, *The reality of VR/AR growth*, 2017.
[29] W. J. Mitchell, K. A. Szerszen Sr, A. S. Lu, P. W. Schermerhorn, M. Scheutz and K. F. MacDorman, "A mismatch in the human realism of face and voice produces an uncanny valley", i-Perception, 2 (2011), pp. 10-12.
[30] M. Mori, "The uncanny valley", Energy, 7 (1970), pp. 33-35.
[31] M. Mori, K. F. MacDorman and N. Kageki, "The uncanny valley [from the field]", IEEE Robotics & Automation Magazine, 19 (2012), pp. 98-100.
[32] Pew, 2017 http://www.pewresearch.org/fact-tank/2017/12/12/nearly-half-of-americans-use-digital-voice-assistants-mostly-on-their-smartphones/.
[33] L. P. Robert, A. R. Denis and Y.-T. C. Hung, "Individual swift trust and knowledge-based trust in face-to-face and virtual team members", Journal of Management Information Systems, 26 (2009), pp. 241-279.
[34] J. B. Rotter, "Interpersonal trust, trustworthiness, and gullibility", American psychologist, 35 (1980), pp. 1.
[35] U. Schultze, "The avatar as sociomaterial entanglement: a performative perspective on identity, agency and world-making in virtual worlds", (2011).
[36] U. Schultze and W. J. Orlikowski, "Research commentary—Virtual worlds: A performative perspective on globally distributed, immersive work", Information Systems Research, 21 (2010), pp. 810-821.
[37] M. Seymour, K. Riemer and J. Kay, *Interactive Realistic Digital Avatars-Revisiting the Uncanny Valley*, *Proceedings of the 50th Hawaii International Conference on System Sciences*, 2017.
[38] M. Seymour, Riemer, K., and Kay, J., "Actors, Avatars and Agents: Potentials and Implications of Natural Face Technology for the creation of Realistic Visual Presence", Journal of association for information systems (2018).
[39] A. Tinwell, M. Grimshaw, D. A. Nabi and A. Williams, "Facial expression of emotion and perception of the Uncanny Valley in virtual characters", Computers in Human Behavior, 27 (2011), pp. 741-749.
[40] L. Wan, K. Crookes, A. Dawel, M. Pidcock, A. Hall and E. McKone, "Face-blind for other-race faces: Individual differences in other-race recognition impairments", Journal of Experimental Psychology: General, 146 (2017), pp. 102.
[41] S. Wang, S. O. Lilienfeld and P. Rochat, "The uncanny valley: Existence and explanations", Review of General Psychology, 19 (2015), pp. 393.
[42] E. Williams, "Experimental comparisons of face-to-face and mediated communication: A review", Psychological Bulletin, 84 (1977), pp. 963.