Facing the Artificial: Understanding Affinity, Trustworthiness, and Preference for More Realistic Digital Humans

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

In recent years, companies have been developing more realistic looking human faces for digital, virtual agents controlled by artificial intelligence (AI). But how do users feel about interacting with such virtual agents? We used a controlled lab experiment to examine users' perceived trustworthiness, affinity, and preference towards a real human travel agent appearing via video (i.e., Skype) as well as in the form of a very human-realistic avatar; half of the participants were (deceptively) told the avatar was a virtual agent controlled by AI while the other half were told the avatar was controlled by the same human travel agent. Results show that participants rated the video human agent more trustworthy, had more affinity for him, and preferred him to both avatar versions. Users who believed the avatar was a virtual agent controlled by AI reported the same level of affinity, trustworthiness, and preferences towards the agent as those who believed it was controlled by a human. Thus, use of a realistic digital avatar lowered affinity, trustworthiness, and preferences, but how the avatar was controlled (by human or machine) had no effect. The conclusion is that improved visual fidelity alone makes a significant positive difference and that users are not averse to advanced AI simulating human presence, some may even be anticipating such an advanced technology.

1. Introduction

In recent years, there has been significant growth in the use of digital, virtual agents controlled by artificial intelligence (AI). Voice-controlled virtual agents are popular in a wide range of consumer products. Nearly half of U.S. adults (46%) say they now use these applications to interact with smartphones and other devices [33]. Apple's agent Siri and Amazon's Alexa[1] are commercial products that blur the line between humans and virtual agents. As a result, the distinction between human agents and virtual agents is growing smaller every year [34].

Newer virtual agents are beginning to take visual form in the online world, with realistic, interactive, fully rendered human faces that cross the "uncanny valley". This theory is widely known for capturing the phenomenon of "eeriness".

A disembodied voice as the representation of the assistant is being expanded to include a more human realistic face with a goal of achieving Realistic Visual Presence (RVP). This is the sensation of human-like presence obtained from interacting with a digital human entity [32]. There has been a steady move towards creating characters and avatars that are more and more visually realistic [33]. The development of RVP is an important area of research as humans are hard wired to respond to human faces in unique and positive ways [20]. Artificial human faces hold great promise for advancing human-computer interaction (HCI) and increasing affinity between humans and their machines [35]. Similarly, few would argue that the misuse of such digital human technology could also lead to an abuse of trust and negative impacts.

We need to understand how users react to new human-like digital entities at the heart of the creation of RVP. So-called cognitive agents with natural faces will likely make their way into real-life contexts. Questions will arise regarding both the ability and desirability of such agents to build relationships with users over time, and the impact that these digital humans will have on our professional and social identities. For example, affinity and trust in digital human entities are important factors that influence whether consumers purchase from online retailers [7].

We distinguish between avatars directed by humans, and agents directed by artificial intelligence. We define a realistic digital avatar (or digital avatar),

[1] Commercial products: see www.apple.com/siri/ & developer.amazon.com/alexa

URI: https://hdl.handle.net/10125/64316
978-0-9981331-3-3

as "an avatar with the realistic, interactive facial representation of the human actor puppeting it" [32]. A visually rich cognitive agent (or digital agent), is defined as an AI-driven entity "visually presented as an interactive, real-time rendered human-like entity, on a screen or in a virtual environment" [32].

Past research suggests that when people believe that digital human entities are driven by a real person, they are seen as engaging [34]. The question thus arises, what if they were driven by a similarly emotionally engaging advanced AI engine, would they be seen as similarly engaging? In prior research users speculated that they would find an advanced digital agent with a realistic human face to be "creepy", "spooky", or "too much" [34], i.e. that the known simulation of advanced human presence would itself be off-putting. If proven true, this would undermine any assumptions that the path to affinity with digital humans can be achieved with improved visuals and simulated realistic interaction. If fully natural realistic responses from an AI were to be inherently unwelcome by users, it would have profound negative implications for the research and practitioner community seeking to build more life-like agents. Such a finding would suggest deliberately stylized human forms may be a better way forward.

In this paper, we compare users' reactions to working with a real human agent using video (similar to Skype) to their reactions to working with a digital human entity controlled by either the same human (digital avatar) or what they believed to be AI (digital agent) when the digital agent was actually puppeted by the same human, with the participants deceived into believing it was controlled by AI in a "Wizard of Oz"-style setting. This study thus asks two questions:

RQ1: Are there differences in user perceptions of trustworthiness, affinity and preferences for human actors and visually realistic digital human entities?

RQ2: Are there differences in user perceptions of trustworthiness, affinity and preferences for digital avatars (controlled by humans) and digital agents (controlled by AI)?

Our results show participants' affinity, trust, and preference were strongly influenced by the visual (Skype or avatar) but not by whether the avatar was controlled by a human or AI. This has both theoretical and practical implications. The first theoretical implication is that reactions were more strongly influenced by subconscious visual aspects of the "person" than by the content of the interaction. We have long theorized that behavior is central to trust-building but our results challenge this by indicating that visual appearance trumps behavior. A second implication is the lack of difference between human and AI control: theory suggests that humans are more capable of benevolence and integrity than machines, yet the AI agent was no less trustworthy. Thus these assumptions are unfounded or benevolence and integrity are unimportant in this context.

The first implication for practice is that improvements to human simulated behaviours is important and will not inherently adversely influence user. The second implication is that to improve responses toward digital humans, work is still needed in developing further their visual appearance, as the same behaviour was found to be less effective when exhibited by an entity with artificial appearance.

The paper is divided in a brief discussion of the theoretical background of this work, our methodology for exploring the research questions, the results, discussion and conclusions.

2. Theoretical Background

In this study, we are interested in three theoretically distinct types of interaction (see Table 1). The first is a human agent who interacts with participants over computer-mediated video; we call this the Video Human (VH) treatment. The second is a human agent who interacts with participants using a digital avatar; we call this the Avatar Human (AH) treatment. The third is an AI-controlled digital agent which presents to participants with the same digital face as the digital avatar; we call this the AI Agent (AA) treatment. We informed the participants in our study that the AA was controlled by IBM Watson (since it recently had been in the news and was known to our participants), even though it was puppeted by the same human as in the other two treatments.

2.1 Prior Research

Digital agents have been the focus of research by many IS scholars. Developing the algorithms that drive digital agents is an important area of research, but it is equally important to understand users' behavior, decision making processes, and attitudes towards those agents [43, 3]. After all, even the best advice provided by a digital agent is useless unless the humans using the agent are willing to accept its advice. This research shows that consumers' beliefs, attitudes, perceptions and behaviors are influenced by the representation or display of the agent or an object, especially if it triggers anthropomorphizing – a in which users ascribe human-like characteristics to the agent [23, 44, 45, 13, 11, 1].

Table 1. The Three Conditions in the Study			
Condition	**Name**	**Definition**	**As Presented in the Study**
Video Human	VH	A human agent interacting via video	A human agent on a Skype-like video call
Avatar Human	AH	A human agent interacting via a digital avatar.	A human agent represented by a digital avatar visually resembling the human agent.
AI Agent	AA	An AI-controlled digital agent entity with no human involvement in the interaction.	An IBM Watson AI-controlled digital agent visually resembling the human agent.

One way to trigger anthropomorphizing is by adding a face to a virtual agent [44, 45]. Digital human entities with more human-like faces are beginning to appear in certain high-end applications and research. For example, BabyX is a virtual agent which presents as a young child [29]. BabyX is based on a self-learning neural brain model. 'She' works using biologically based computational behavioral models that determine 'her' baby-like behavior [12]. BabyX uses a psychobiological modeling framework called 'Brain Language' (BL) to create an autonomous virtual agent with a human realistic face and facial motion. BabyX's autonomous expressive behavior is driven by various neural-system models based on affective and cognitive neuroscience theories. As a result, she resembles an 'unscripted agent', in that her responses are constructed in the moment, using the latest theories in childhood neuroscience [29].

The BL was designed to support a wide range of computational neuroscience models, as documented in [38]. These models are integrated into a consistent system of responses that range from simple leaky integrators to spiking neurons to mean field models to self-organizing maps. The BL is deployed within the infrastructure designed to support it, using larger neural networks (such as convolutional networks like those used in deep learning and recurrent networks). This allows a BL coded agent to simulated emotional responses that can provide plausible facsimiles of emotional intelligence, without actually replicating emotional or rational thinking.

BabyX is one of the most well-known human-like digital agents. Other highly realistic digital agents that build on BabyX, include Rachel, Nadia and Roman, each of which are driven by forms of AI that engages in natural language conversation, and emotional intelligence [33]. They do this using BL coupled with natural language understanding (NLU) and natural language generation (NGL), in a limited domain space.

Other examples of agents with human like facial representations include the growing class of virtual influencers such as Lil Miquela, Blawko and Bermuda [36]. Although these digital humans are most often not interactive.

Research has shown that when presented with two human-controlled digital avatars, one highly realistic and the other a cartoon character, people rated the realistic avatar as more trustworthy, had more affinity for it and equally preferred it as a virtual agent. However, the same study questioned if our acceptance of these highly realistic human face representations may be moderated by who is perceived to be controlling the entity, a human or an AI [34].

2.2. Self-categorization Theory

Self-categorization theory argues that individuals automatically categorize the other either as in-group or out-group members during social interactions [39]. They tend to trust and favor in-group members over out-group members. Attitude, perceptions, and emotions are easier to be shared among in-group members than out-group members [39].

This same in-group/out-group process also applies to digital characters. Prior research found that individuals playing video games are more likely to attribute in-group digital characters with the same emotional state than out-group ones [4]. Likewise, they also tended to mimic the emotion from their in-group digital characters. Thus, individuals are prone to anthropomorphizing digital entities with human-like features [24].

Research suggests that users may see a human-controlled digital avatar either as a direct extension of the user controlling it or as a separate and distinct entity [31]. 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 them. This is the same process that occurs when we see ventriloquist interacting with a dummy; we know the dummy is being controlled by the ventriloquist, but part of us ascribes the dummy with agency separate from the ventriloquist.

We theorize that self-categorization theory may be related to affinity with the digital humans and affected by their perception of identity. This would extend the notion of affinity with a digital human to be seen as the key factor extending beyond aesthetics and directly influencing trustworthiness.

2.3. Affinity and the Uncanny Valley

The almost 50-year-old Uncanny Valley theory [22] 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 close to being realistic, at which point affinity drops dramatically because a semi-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.

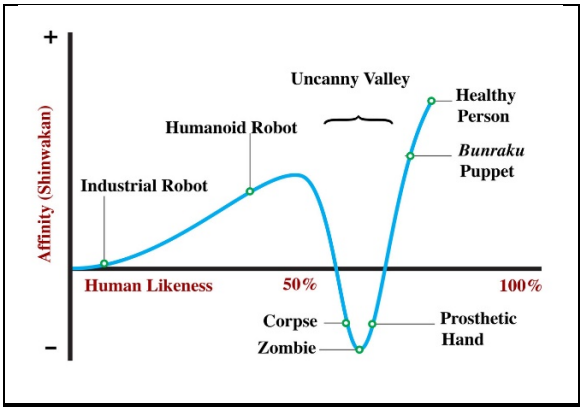

Figure 1. The Uncanny Valley

The Uncanny Valley uses the concept of "affinity", which comes from an original Japanese word, Shinwakan (親和感), and is open to interpretation as to how it is translated into English. "Affinity" has emerged as the preferred translation [42, 22]. Affinity is an indicator of whether an avatar is in the Uncanny Valley.

The cause(s) of the Uncanny Valley are not clear, but there are many different theories (see [42] 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 [22, 30]. 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 [42]. 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) [41, 9]. 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 [22]. We perceive almost-human avatars to be humans exhibiting a psychopathic personality disorder [37]. 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 [33], so, we have little understanding of how users perceive interacting avatars, especially those with highly realistic faces. A digital avatar with a realistic face will be closer to a real human, but still visually different. Users will still have greater affinity for a human video than a digital avatar which looks slightly less human-like, regardless of who or what is controlling it. Thus:

> ***Hypothesis 1a.*** *Individuals will have greater affinity for a human agent using video (VH) than either a digital avatar or digital agent (AH and AA).*

Knowing the digital avatar is controlled by a human, individuals will be consciously aware that the avatar is just an extension or different form of another human being. The consciousness, attitude, perception, and behavior of the avatar can be explained using rules applicable to humans. Individuals will be more likely to categorize a digital avatar they believe to be controlled by a real person as an in-group member (the group of human) than a digital agent they believe is controlled by an AI. Therefore, we theorize,

> ***Hypothesis 1b.*** *Individuals will have greater affinity for a digital avatar controlled by a human*

(AH), than a digital agent controlled by artificial intelligence (AA).

2.4. 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 [17]. Trustworthiness is an assessment of whether another person or thing is worthy of trust [17]. Trust is between people [17], but also applies to information systems [16, 40, 3, 12].

Mayer, et al. [17] 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 [14, 17]. The trustor's disposition to trust is independent of the trustee; it is a "generalized attitude" learned from experiences of fulfilled and unfulfilled promises [22, 27, 28], and varies from person to person.

The other three elements of trust are based on the trustor's assessment of the trustee [10, 17, 26]. 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 [8]. 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 [8]. 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 humans are more likely to be perceived as having integrity and benevolence than non-human agents controlled by AI. Because integrity and benevolence affect trustworthiness, we theorize that human agents will be perceived as more trustworthy than either digital avatar or agent. Thus:

Hypothesis 2a. *Individuals will ascribe greater trustworthiness to a human agent using video (VH) than to a digital human entity, digital avatar or agent (AA and AH).*

People trust in-group members more than out group members. Similar to affinity, we believe people will more likely to categorize a human-controlled avatar as an in-group than AI-controlled agents. Therefore:

Hypothesis 2b. *Individuals will ascribe greater trustworthiness to a digital avatar controlled by a human (AH) than a digital agent controlled by artificial intelligence (AA).*

2.5. User Preferences

Affinity and trustworthiness are two important characteristics of virtual agents [7]. Affinity has often been linked to increased preferences for interaction with avatars and web sites in general [5, 7, 15]. Likewise, trustworthiness is an important factor influencing both interpersonal preferences and preferences for websites – and increased interactions with both [8, 18]. We argued above that interaction with a human would induce greater affinity (Hypothesis 1a) and greater trustworthiness (Hypothesis 2a) than interaction with a digital entity of any sort. Taken together, we theorize that humans should be preferred as agents. Thus:

Hypothesis 3a. *Participants will prefer a human agent using video (VH) to either digital avatar or agent (AA and AH).*

Avatars with human "mind" behind them will be more likely categorized as a closer in group member. People favor in-group members over out group members, such attitude can be translated as higher level of preferences in a self-report format. Therefore,

Hypothesis 3b. *Individuals will prefer a digital avatar controlled by a human (AH), to a digital agent controlled by artificial intelligence (AA).*

3. Method

We conducted a 2x2 repeated measures laboratory experiment to test the hypotheses.

3.1. Participants

67 undergraduate students at a large university in Australia participated in the experiments. Five participants were excluded due to technical failures,

resulting in a total sample of 62. About 60% of the participants identified as female, 38% as male and one identified as other. The average age was 23.6. Subjects' participation was voluntary; each participant received a free movie ticket.

3.2. Task

The task participants performed was to use a travel agent to get a quote for the airfare portion of an overseas trip. Participants performed the same task twice with different destinations: a trip from Sydney to Los Angles, and a trip from Sydney to London. The same script, with questions in the same order, was used for both tasks in all treatments (e.g., first the travel dates, then class of service, and so on).

3.3. Treatments

The within-subjects factor of the repeated measures design was the agent type (human agent on video (VH) vs. digital human entity) and the between-subject factor was the type of digital entity, human-controlled digital avatar (AH) vs. AI controlled digital agent (AA). Thus every subject received the VH treatment, and then either one of the AH or AA treatments. Treatment order (human or digital first), task order (Los Angles or London first), and the type of digital entity (AH or AA), were randomly assigned.

The VH treatment was implemented using a video application similar to Skype. The background and environment was the same for all treatments. A neutral nighttime background was placed behind the real actor and the digital entities so their settings would not be a factor. The video human agent and digital agent wore the same style of clothing and were side lit (but for technical reason from opposite sides). Their audio was identical in quality and reproduction.

Participants were deceived into believing the AA treatment used AI when in fact it did not. In both AH and AA treatments they interacted with a digital avatar controlled by the same person, who was also the human agent in the VH treatment. In fact the only treatment that participants received was a briefing that either 1) rightly informed them that they were about to interact with a real human (AH) via a digital avatar, or 2) deceptively told them that they were about to interact with a digital agent controlled by AI (AA), specifically IBM Watson.

The digital avatar used in the AH and AA treatments was developed based on the image of the same person as in the VH treatment (see Figure 2). It needs to be noted that the creation of the 3D natural digital human face of the avatar, to be puppeted in real time, took considerable effort. The highly realistic digital avatar was based on a custom-built, advanced facial tracking and live animation system that responded in real-time and without noticeable latency. The avatar was developed by a team comprised of the researchers, professional video game animators working in well-known technology firms, and professional technical directors working for well-known real time gaming companies. It matches appropriately to industry standards. While not indistinguishable from a real person, this model had previously been used and been found to be trustworthy in such a context [34].

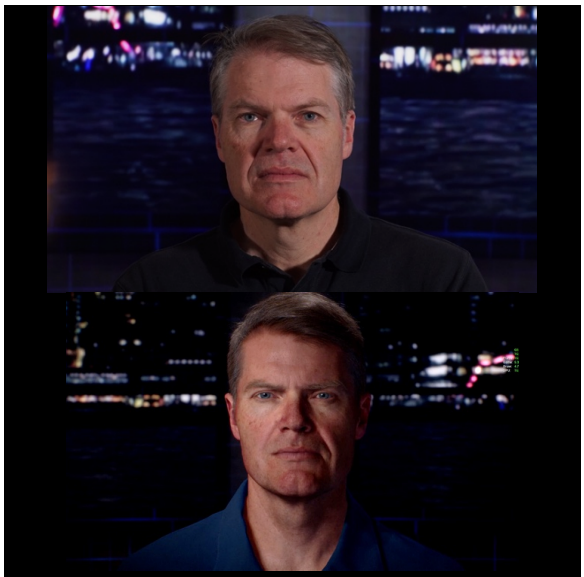

Figure 2. The video human agent (VH treatment) above and RVP Avatar agent (AH and AA treatments) below.
Photographs courtesy of the authors

3.4. Measurements

Measurements were adopted from prior research and modified to assess the constructs of interest in this study. The measures used in this study are summarized in Table 1. All affinity and trustworthiness items were measured using a 7-point Likert scale with 1 representing strongly disagree and 7 indicating strongly agree. Measurements for both constructs were reliable with Cronbach's alpha higher than 0.70. Participants were asked to indicate their preference for a video human agent (VH) or the digital avatar/agent (AH or AA). The question was 7-points anchored at zero in the middle with 3s on the ends for strongly prefer video agent and strongly prefer digital agent. We converted the responses so that positive numbers indicated a preference for the video human (VH).

3.5. Procedures

The experimental procedure was the same for all participants. After arriving at the laboratory, they filled out a consent form, received instructions about the procedure, including an explanation for the treatments, and watched a short explanatory video. They then performed one of the two travel tasks and completed a short survey. This was repeated with the second task and a second survey. They then moved to a second room and had a recorded qualitative interview, were debriefed, and received payment in the form of a cinema voucher.

Table 1. Measurement Items for Trustworthiness and Affinity

Construct	Reliability (Cronbach's alpha)	Item
Trust-worthiness [25]	0.85 Human (VH) 0.82 Avatar (AH and AA)	I think the agent is trustworthy.
		I think the agent is honest.
		I think the agent is dependable.
		I think the agent is competent.
Affinity [2]	0.71 Human (VH) 0.79 Avatar (AH and AA)	The agent is lifelike.
		The agent is friendly.
		I felt affinity with the agent.

4. Results

Statistical analyses on trustworthiness, affinity, preferences were completed in SPSS 23 using repeated-measures general linear model (GLM). A power analysis shows that our study has sufficient power (.972) to detect medium effect sizes for within subject tests and for the interaction between within and between subject treatments.

Table 2 presents treatment means Participants reported higher level of affinity for the video human agent (VH) than the two RVP avatars ($F(1, 58)=24.678$, $p<0.001$); however, there was no statistical difference between human-controlled avatar (AH) or AI agent (AA) ($F(1, 58)=1.348$, $p=0.250$). The order of the tasks (London or Los Angeles) had no effects, neither within subject ($F(1, 58)=0.263$, $p=0.610$) nor between subjects ($F(1, 58)=0.606$, $p=0.439$). The order of the treatments has no within subject effects ($F(1, 58)=2.219$, $p=0.142$) but between subject effect ($F(1, 58)=14.326$, $p<0.001$). H1a is supported but H1b is not.

Participants had more trust in the video human (VH) than in the two digital treatments ($F(1, 58)=22.229$, $p<0.001$). There is no significant differences on trust between participants who believed that the avatar was controlled by human (HA) and those who believed that they interacted with an AI-controlled agent (AA) ($F(1, 58)=1.270$, $p=0.264$). The order of the tasks has no within subject effect ($F(1, 58)=0.310$, $p=0.580$) or between subject effect ($F(1, 58)=0.979$, $p=0.327$). The order of treatments had no within subject effect ($F(1, 58)=0.020$, $p=0.823$) but between subjects effect ($F(1, 58)=6.331$, $p=0.015$). H2a is supported but H2b is not.

We used t-tests to compare the preferences to neutral (i.e., =0). We found participants in both treatments to prefer the video Human (VH) to the digital avatar/agent (AH: $t(32)=5.43$, $p<.001$; and AA: $t(28)=6.40$, $p<.001$). There were no significant differences in preferences towards the digital entity when participants believed they interacted with an avatar controlled by a human (AH) and those believed they interacted with a digital agent controlled by AI (AA) ($F(1, 58)=0.318$, $p=0.575$). Neither treatment order ($F(1, 58)=0.735$, $p=0.395$) nor task destitution city ($F(1, 58)=0.814$, $p=0.371$) had any effects on preference. H3a is supported but H3b is not.

Table 2. Means and Standard Deviations

	Affinity	
	Mean	Std.
Video Human (VH)	5.941	0.831
Avatar Human (AH)	5.273	1.298
Avatar AI (AA)	5.207	1.033
	Trustworthiness	
	Mean	Std.
Video Human (VH)	6.113	0.842
Avatar Human (AH)	5.659	1.053
Avatar AI (AA)	5.491	1.034
	Preference*	
	Mean	Std.
Video Human (VH)	na	na
Avatar Human (AH)	1.700	1.795
Avatar AI (AA)	1.930	1.624

* Positive numbers indicate a preference for the Video Human (VH) over the AH or AA avatar

5. Discussion

Our results show that participants perceived the video human agent (VH) to be more trustworthy and had more affinity for him than either the digital avatar

(AH) or the digital agent (AA). Likewise, participants preferred the human agent (VH) to either the digital avatar (AH) or agent (AA).

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 [19]. Perhaps it is not surprising then, that our participants preferred human agents using video (e.g., Skype) to both kinds of digital face representation, as indicated in the Uncanny Valley Theory.

More interestingly, there were no differences in affinity, trustworthiness or preference between the digital avatar controlled by a human (AH) and the digital agent controlled by AI (AA). We would have expected that people develop aversion to a purely digital entity once they know that it is controlled by AI, and as a result trust it less, and categorize it as far less similar to themselves. However, our results show that once the AI achieves the same level of realism in behavior as human controlled interaction, they are likely to be perceived as the same as humans, even though our participants believed they were AI controlled digital agents.

As a reminder, the digital face utilized was a technology artifact designed to induce a perception of humanness and controlled in both instances by a human. We instructed half of our participants that the face was puppeted by the same person present in the video conversation yet deceived the other half into believing that it was controlled by an AI. Hence, since participants were consciously aware of who or what was controlling the digital face, we reasoned that this knowledge would make a difference in that the avatar controlled by a human would likely be perceived as an in-group member and therefore be more trustworthy than a digital, non-human agent controlled by AI. Based on prior research, we reasoned that an AI with human-like abilities would make our participants less comfortable than the avatar controlled by a human. However, our results did not support this.

Why then were our participants not thrown off by a fully digital entity that exhibits traits of general intelligence and a level of conversational proficiency out of reach of the current state of technology? One clue was given in our exit interviews. After completing the study, participants were interviewed about their experiences. One question was to make sure that they had accurately understood the treatment they were in, whether AH or AA. About 90% of the participants (56 out of 62) correctly understood which treatment they had received. However, six of the AH participants, who were clearly briefed that the digital face was driven by a real human, nonetheless reported that they believed (wrongly) that they had interacted with an agent controlled by AI. We compared the affinity, trustworthiness, and preferences of these six participants to the others in the AH treatment and found no significant or meaningful differences.

We reason that those six participants might have assumed a unity in the entity that they interacted with; because the face was artificial, they extrapolated subconsciously that the controlling entity must also be artificial. Alternatively, they might not have been able to envision how a digital face could be puppeted by a human. In either case, what is significant about this observation is how ready the cohort of participants was to believe in the existence of general AI, at a level that is beyond current capability. We argue that it is this deep-rooted belief in the power of technology that is at the heart of the above finding; namely that the AA treatment did not lead to negative perceptions that would have it rendered different to the AH treatment.

This observation opens up an interesting new stream of research to investigate how personal, and collective, attitudes towards technological progress in general and AI in particular, might influence people's readiness for engagement with new digital human technologies, in different ways. It may be that young adults who have grown up in a technology-rich world might have different responses than older adults who remember a time when technology was not so ubiquitous and powerful. We would expect differences to exist among different age groups and among people with different technology backgrounds. We propose future studies should both investigate attitudes towards AI and experimental research on digital humans taking into account such potential differences.

Studies like ours suffer from the common limitations of lab experiments with undergraduate students working on artificial tasks. Student samples are considered to be an appropriate for testing theories about phenomena that expected to hold true across the general population [6], and there is evidence that students have an equal exposure to avatars as general population in real life. The task was artificial in that participants were not required to actually pay for the inquired trip to the destination cities; there were no consequences to planning the trips.

Our study has several implications. First, users have more affinity for and trust in humans interacting over video and prefer them to digital avatars controlled by humans (or by AI). For services provided via interpersonal interaction, people still prefer real humans. Companies should still consider using human agents for such tasks. Second, participants in our study, a young and well-educated cohort of people, did not prefer or trust avatars controlled by humans more than digital agents controlled by AI, as long as AI

controlled agents are implemented on the same level of perceived intelligence. However, this may be tempered to some extent by the application. Companies who plan to use AI controlled agents should focus on creating digital representations that "feel as if" they are driven by real humans. The design and the findings of this research can be used by real companies and provide opportunities to observe consumers' behaviors in real life and produce generalizable findings.

6. Conclusion

The development and use of digital avatars and AI-based digital agents has been the interest of multiple disciplines and industry alike. Our goal was to contribute to the body of knowledge on this topic by directly comparing individuals' trustworthiness, affinity, and preference for human agents and digital counterparts (whether human controlled or AI-based). This study provides some initial evidence in understanding individual perceptions of digital avatars and agents.

Understanding individual perceptions is difficult and complex. What we found was that visual aspects were more important than the behaviour in driving user perception. One implication is that although research on AI agents to improve their behaviour is important, the development of human-realistic faces is also important. If we expect users to trust and have affinity for the AI agents they work with, visual appearance is critical. Even the highest performing AI agent will not engender high levels of trust, affinity, and preference to use unless the AI agent looks the part. Development of appearance, especially of the digital face, is essential because human perception is not only rational and influenced by behaviour, but is strongly – perhaps even more strongly – influenced by visual appearance.

The research shows that in terms of the visuals alone, while the level of realism between the artificial and the actual seems subtle and involving only small differences, those differences are important, and additional improvements in realism would have a significant effect.

We expect our research to contribute to the practical challenges of producing more useful digital humans, as practitioners spend millions of dollars to push digital agent technologies forward, while companies make deployment decisions, and users begin to encounter such entities "in the wild".

Our research is just one step in the process of refining our understanding of human perception, attitude, behavior, and decision making in this context.

7. References

[1] H. Abdul Rahman, J.-S. Choe and J. Park, "Offer Strategy Model of Integrative Negotiation for Automated Negotiation Agent: Multiple Equivalent Simultaneous Offers and Argumentation-based Negotiation", (2017).

[2] C. Bartneck, T. Kanda, H. Ishiguro and N. Hagita, Is the uncanny valley an uncanny cliff?, Robot and Human interactive Communication, 2007. RO-MAN 2007. The 16th IEEE International Symposium on, IEEE, 2007, pp. 368-373.

[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] A. Besmann and K. Rios, "Pals in power armor: Attribution of human-like emotions to video game characters in an ingroup/outgroup situation", Cyberpsychology, Behavior, and Social Networking, 15 (2012), pp. 441-443.

[5] 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.

[6] D. Compeau, B. Marcolin, H. Kelley and C. Higgins, "Generalizability of information systems research using student subjects - a reflection on our practices and recommendations for future research", Information Systems Research, 23 (2012), pp. 1093-1109.

[7] 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.

[8] 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.

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

[10] S. L. Jarvenpaa and D. E. Leidner, "Communication and trust in global virtual teams", Journal of Computer-Mediated Communication, 3 (1998),

[11] D. Kim, K. Park, Y. Park, J. Ju and J.-H. Ahn, Alexa, Tell Me More: The Effect of Advertisements on Memory Accuracy from Smart Speakers, Proceedings of the Pacific Asia Conference on Information Systems (PACIS), 2018.

[12] 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.

[13] K. Letheren, S. Lee, D.-H. Kwak and K. Ramamurthy, "Effects of Gendered Anthropomorphism and Image Appeal on Moral Norms in the Context of Charity Website Design", (2018).

[14] R. Lewicki and B. Bunker, Conflict, cooperation and justice, chapter Trust in relationships: a model of trust development and decline, Jossey-Bass, 1995.

[15] 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.
[16] 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.
[17] 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.
[18] 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.
[19] 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.
[20] M. Meng, T. Cherian, G. Singal, & P. Sinha (2012). Lateralization of face processing in the human brain. *Proceedings of the Royal Society B: Biological Sciences*, *279*(1735), 2052–2061. http://doi.org/10.1098/rspb.2011.1784
[21] 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.
[22] M. Mori, K. F. MacDorman and N. Kageki, "The uncanny valley [from the field]", IEEE Robotics & Automation Magazine, 19 (2012), pp. 98-100.
[23] R. Ohanian, "Construction and validation of a scale to measure celebrity endorsers' perceived expertise, trustworthiness, and attractiveness", Journal of advertising, 19 (1990), pp. 39-52.
[24] Proudfoot, D. (2011) 'Anthropomorphism and AI: Turings much misunderstood imitation game', *Artificial Intelligence*. Elsevier B.V., 175(5–6), pp. 950–957. doi: 10.1016/j.artint.2011.01.006.
[25] L. Qiu and I. Benbasat, "Evaluating anthropomorphic product recommendation agents: A social relationship perspective to designing information systems", Journal of Management Information Systems, 25 (2009), pp. 145-182.
[26] 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.
[27] J. B. Rotter, "Interpersonal trust, trustworthiness, and gullibility", American psychologist, 35 (1980), pp. 1.
[28] J. B. Rotter, "A new scale for the measurement of interpersonal trust", Journal of personality, 35 (1967), pp. 651-665.
[29] M. Sagar, M. Seymour and A. Henderson, "Creating connection with autonomous facial animation", Commun. ACM, 59 (2016), pp. 82-91.
[30] A. P. Saygin, T. Chaminade, H. Ishiguro, J. Driver and C. Frith, "The thing that should not be: predictive coding and the uncanny valley in perceiving human and humanoid robot actions", Social cognitive and affective neuroscience, 7 (2011), pp. 413-422.
[31] U. Schultze, "The avatar as sociomaterial entanglement: a performative perspective on identity, agency and world-making in virtual worlds", (2011).
[32] M. Seymour, K. Riemer and J. Kay, "Actors, Avatars and Agents: Potentials and Implications of Natural Face Technology for the Creation of Realistic Visual Presence", Journal of the Association for Information Systems, 19 (2018), pp. 953-981.
[33] 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.
[34] M. Seymour, L. Yuan, A. Dennis and K. Riemer, Crossing the Uncanny Valley? Understanding Affinity, Trustworthiness, and Preference for More Realistic Virtual Humans in Immersive Environments, Proceedings of the 52nd Hawaii International Conference on System Sciences, 2019.
[35] L. Sproull, M. Subramai, S. Kiesler, J.H. Walker & K. Waters (1996). When the interface is a face. *Human-Computer Interaction*, *11*(2), 97–124.
[36] Tiffany, K. (2019) *Lil Miquela and the virtual influencer hype, explained*, *Vox.com*. Available at: https://www.vox.com/the-goods/2019/6/3/18647626/instagram-virtual-influencers-lil-miquela-ai-startups (Accessed: 1 September 2019).
[37] 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.
[38] T. Trappenberg, *Fundamentals of Computational Neuroscience.* Oxford University Press, New York, 2010.
[39] J. C. Turner, M. A. Hogg, P. J. Oakes, S. D. Reicher and M. S. Wetherell, Rediscovering the social group: A self-categorization theory, Basil Blackwell, 1987.
[40] A. Vance, C. Elie-Dit-Cosaque and D. W. Straub, "Examining trust in information technology artifacts: the effects of system quality and culture", Journal of management information systems, 24 (2008), pp. 73-100.
[41] G. T. Viki, L. Winchester, L. Titshall, T. Chisango, A. Pina and R. Russell, "Beyond secondary emotions: The infrahumanization of outgroups using human–related and animal–related words", Social Cognition, 24 (2006), pp. 753-775.
[42] S. Wang, S. O. Lilienfeld and P. Rochat, "The uncanny valley: Existence and explanations", Review of General Psychology, 19 (2015), pp. 393.
[43] B. Xiao and I. Benbasat, "E-commerce product recommendation agents: use, characteristics, and impact", MIS quarterly, 31 (2007), pp. 137-209.
[44] L. Yuan and A. Dennis, Interacting Like Humans? Understanding the Effect of Anthropomorphism on Consumer's Willingness to Pay in Online Auctions, Proceedings of the 50th Hawaii International

Conference on System Sciences, 2017.
[45] L. Yuan, A. Dennis and R. Potter, "Interacting Like Humans? Understanding the Neurophysiological Processes of Anthropomorphism and Consumer's Willingness to Pay in Online Auctions", (2016).