Evaluating Realism for Virtual Supporting Characters

James Skorupski*, Josh McCoy*, Catherine Zanbaka†, Kathy Ryall†, Michael Mateas*

*University of California, Santa Cruz
1156 High Street
Santa Cruz, CA 95064
{jskorups, mccoyjo, michaelm}
@soe.ucsc.edu

†BAE Systems
6 New England Executive Park
Burlington, MA 01803
{catherine.zanbaka, kathy.ryall}
@baesystems.com

ABSTRACT

Convincing and compelling virtual environments that are populated with rich characters demand consistent, nuanced, and realistic behavior that is integrated in the surrounding environment. Inadequate and unscalable traditional Artificial Intelligence based approaches have made it impractical to apply this level of character depth and detail to large environments. We introduce the concept of *Supporting Character Realism (SCR)* for virtual avatars by identifying the capabilities of agents which have the ability to work in tandem with traditional "main character realism" approaches by demonstrating consistent and nuanced behaviors that blend into the surrounding background environment. Next, we propose several metrics for evaluating agents attempting to achieve this level of realism and test our proposed metrics in a rich social interaction experiment placed in a virtual bar amidst a variety of human and computer controlled patrons. Finally, we gauge the performance of a set of both traditional scripted bots and prototype AI-driven agents designed to target our concept of *SCR*. Our results show that *SCR* is not only a distinguishable and measurable metric of agent realism, but also a technically achievable goal within the reach of modern AI techniques.

Author Keywords

Avatars, Realism, Virtual worlds, Artificial Intelligence.

ACM Classification Keywords

H5.m. Information interfaces and presentation (e.g., HCI): Miscellaneous.

General Terms

Design, Experimentation, Measurement

INTRODUCTION

Interactions involving many individuals mediated by technology, specifically in virtual environment settings, are now a commonplace occurrence. These virtual environments vary widely in purpose and include social networking and communication [7, 15]; virtual story worlds for entertainment and game play purposes, such as massively-multiplayer online roleplaying games [20], competitive multiplayer games that include thousands of audience members [8] and education and training [14]. Experiential training focused on social skills and human interaction, such as medical personnel practicing bedside manner, learning interviewing techniques, or practicing police work, may require the use of computer controlled non-player characters (NPCs) to populate the space, as having human trainers control all characters doesn't scale for broad deployment nor provide desired repeatability. The study reported here is motivated by the need for NPCs that trainees respond to as if they were human controlled avatars in the virtual world. This requires defining behavioral baselines for human-to-human interaction in virtual worlds, and comparing human-NPC interaction against this baseline.

Human-to-human virtual world interactions have been much more heavily studied [3,5,12,13,24,27] than interactions between mixed groups of humans and NPCs in virtual worlds. In defining "successful" NPCs, there is often an implicit recourse to realism, where a "realistic" NPC is one that is more or less indistinguishable from human. But "realism" is a complex concept that must first be teased apart before we can hope to measure it. Here we split realism into three distinct categories: believability, Turning Test realism and supporting character realism.

In media such as video games, are often defined as *believable*, a term borrowed from animation and the character arts. Here, a successful NPC is not one that fools the player into thinking they are controlled by a human, but rather one in which the player willingly suspends disbelief [9]. Thus a believable NPC creates a consistent and compelling illusion of life [1], one that a human participant willingly accepts as a representation, without ever thinking about (or caring) whether the character is controlled by a human being. While believability can provide rubrics for the creation of NPCs, some training applications will require trainees to treat NPCs as if they were human, requiring NPCs that don't immediately highlight their representational nature.

In contrast, *Turing Test realism* assumes the adversarial conditions of the Turing Test [21], in which a human participant actively probes to expose a computer masquerading as a human being. In this strongest sense of "realism", an NPC succeeds if it actively fools a participant into thinking it is human in the face of skeptical and probing interactions. This would likely require a complete artificial intelligence model of a human being, making this sense of realism useless for building compelling and effective NPCs now or even potentially in the long term.

We define *supporting character realism* (henceforth SCR) as a new category for determining NPC success. A supporting character is one which engages in background activity and light interaction with the player. A supporting character isn't intended to engage in long-term complex behavior with the player, but rather to provide a sense of realistic human activity around the player. Distinct from believable characters, which the player knows are not real but willingly suspends disbelief, and Turing Test realism, in which the player actively probes to determine if the character is computer or human controlled, with SCR the player doesn't think to question whether a character is human or computer controlled. In a mixed environment of human avatars and SCR agents, in the context of background activity and lightweight interaction with human-controlled avatars, an SCR agent blends in seamlessly. In the context of virtual world training environments, SCR is useful where many characters are needed, such as an urban street scene. If it is possible to build AI characters that achieve SCR, then human trainers would only need to control characters with which the trainee has complex and detailed interactions, while SCR agents fill out the rest of the space. Trainers may need to dynamically switch control as a trainee begins detailed interactions with a formerly background character. But with SCR a trainer would be able to control a large number of characters in the training environment.

To determine whether it is possible to achieve SCR, we developed AI-controlled prototype avatars and tested them in the context of a social scenario implemented in the Second Life[1] (SL) virtual world. The social scenario, set in a bar, is inhabited by a mixture of human controlled avatars, AI-controlled avatars (henceforth "SCR avatars"), and bots, where the bots are implemented using standard Second Life scripting techniques. We first established baseline metrics by which to compare how participants relate to avatars in the virtual world. Then, with human, bot and SCR avatars all interacting and role-playing bar denizens, participants were asked to engage in simple social interactions tasks. If we are successful in implementing SCR, the SCR avatars should lie between bots and human-controlled avatars along the various interaction metrics. And indeed, analysis on the gathered data indicates that SCR avatars are much closer to humans than to bots in the major of SCR metrics.

BACKGROUND

We build on prior research in our definitions of SCR behavior, quantitative and qualitative metrics for characterizing avatar interactions, and in our technical infrastructure and approach for integrating external AI-controlled characters into SL. We used SL as the virtual world environment for exploring SCR. In addition to boasting a substantial player population, SL offers a virtual environment with research-friendly capabilities that include computer controlled avatars, data recording and very customizable environments. Varied research topics including education [3], simulating fire accidents [6], and modes of human-computer interaction [19] have been conducted in Second Life.

SL and other virtual worlds have been used as a methodological tool to study human interactions with both other human and AI controlled avatars [4]. Studies in the social behavior of players, such as accrual of social capital [13], gender and sex practices in virtual worlds [5], social affordances of players [27], decoupling of non-human behavior between players [6], and effects of gender differences between a human and their avatar [26] provide particularly useful tools for both defining the behavior of SCR avatars and developing metrics for characterizing avatar interactions.

We borrow some of our spatial measures from Friedman et al [10]. This work makes use of bots written in the SL scripting language that explore and find objects of interest. The bots recorded information about player's special responses to other avatars and their proximity in dyadic interactions. Given the additional capabilities of our SCR avatars vs. the rather simple behavior supported by the SL scripting language, our study supplements these spatial metrics with metrics such as gesture frequency, avatar facing, and a questionnaire over co-presence.

How social behavior changes over time and expressions of personality in virtual worlds were analyzed in a longitudinal study of a group of around 80 students enrolled in a class about SL [12, 25]. A framework to gather avatar-related information from SL over extended periods of time [24] was used in conjunction with qualitative measures to perform both analyses. Social involvement was shown to increase over time through similar increases in metrics of number of friends, groups, and time spent in populated areas. The subjects' activity and exploration was shown to decrease over time with increases to low-energy actions and time spent in their 3 favorite regions while the metrics of unique regions visited, teleports used and high-energy actions decreased. Qualitative data in the form of weekly questionnaires demonstrated that the distribution of types of activities the students engaged in stabilized over time and was consistent with how time is spent in the real world. Expressions of personality were studied using the measures of a 50-item scale measuring the Big Five factor structure, and avatar-related metrics, which include stance, frequency of logging in, and nearby avatars, and linguistic measures. These behavioral metrics were shown to have high rank-order and low absolute stabilities while very little stability was seen in the linguistic measures. Correlations of conscientiousness and emotional stability with the behavioral and linguistic measures were found. Although

[1] http://www.secondlife.com

our methodology to evaluate SCR is similar, the salient portions of the collected data are different as SCR is evaluated in a single environment with a focus on dyadic interactions between the participant and other avatars. Additionally, since we include human, bot and SCR avatars, different methods of data analysis are necessary to determine the trends of participant's interactions with respect to avatar control type.

Weitnauer [22] makes use of a similar AI architecture to the one we employ to create a proof-of-concept implementation of an AI avatar with more capability than standard SL bots. We make use of similar technical infrastructure in using the the *libOpenMetaverse*[2] framework to connect our AI system and data reporting framework to SL. Max's work was a technical proof of concept, and thus was never evaluated with respect to human interaction.

METHODS

In order to experimentally operationalize SCR, we constructed a social environment that encourages constant interaction and is filled with distinguishable social roles and personalities that would be recognizable to our participants, with supporting characters being played by a mixture of humans, bots and SCR avatars. With these goals in mind, we chose a virtual bar for our environment, defining 5 roles in this setting. Our first experiment, an observational study with humans and simple computer-controlled scripted bots, was constructed to establish the metrics for measuring and differentiating between the actions and interactions of human and computer controlled avatars, and to provide behavioral targets for the SCR avatars based on how our confederates played their roles

Scenario and Roles

In our experiments, a single participant is introduced to our virtual bar filled with both human-controlled avatars (henceforth referred to as "confederates"), and computer-controlled avatars, which are either traditional scripted bots or prototype SCR avatars, both described in following sections. The participant takes the role of a student entering the bar for the first time. In the observational study human confederates and bots performed the following social roles commonly found in bars:

Regular: The veteran of the nightclub/bar. The regulars have their frequented hangout locations and know all of the rules of form of the bar. If someone breaks those rules, the regulars are the first to correct the violators.

Spring Breaker: This gregarious personality is always doing something; dancing, running to the bar for more drinks, bringing others into dance groups, running to up to strangers and getting a groove on: the life of the party.

Wallflower: In social situations, a wallflower is a slang term used to describe shy or unpopular individuals who do not socialize or participate in activities at social events. It is most often used to describe someone who stays close to a wall and out of the main area of social activity.

Waiter/Waitress: As the service staff of the bar, the waiter or waitress caters to the patrons of the bar by serving drinks, and attends to the bar and the various tables around the dance floor.

Generic: Our generic personality embodies a passive but friendly patron that blends into the general activity of the bar. He or she hangs around in various places around the dance floor with a drink in hand, observing the activity around and reacting in kind to dancing and positive gestures from other patrons.

Participants were tasked with various social interactions with patrons of our virtual bar, with the intention of them interacting with every avatar in the virtual bar. These tasks included: getting to know other avatars at the bar, finding the shyest/most outgoing/most popular patron in the bar, buying somebody a drink, hanging out with various social groups, and getting an avatar to buy the participant a drink.

Non-linguistic Interaction

In our experiments, we chose to exclude the voice and chat capabilities of the SL platform, and limit interaction between all avatars in our scenario to physical movement, gestures, dancing, sitting/standing, and giving/receiving drink objects. The goal of this work is to explore SCR, which would typically not include the socially demanding, subtle, and interrogative interactions of chat. While future work will focus on limited supporting character language capabilities, we focused here on physical interaction and performance. In order to establish a credible reason for this limited interaction, our virtual bar environment was constantly bombarded with loud dance music, like that of many real-world bars or clubs.

[2] http://lib.openmetaverse.org

Figure 1: Our test setting consisted of a furnished virtual bar, complete with loud music.

Second Life (SL)

We used SL as our virtual world platform, due to the platform's existing technical support for networked multi-user interaction, expressive avatar gestures, scripted data logging and computer control of avatars. Additionally, SL's large virtual world has an extensive market of virtual assets, including additional avatar gestures, avatar clothing, buildings, furniture and props that allowed us to easily shape our virtual bar into a more plausible environment filled with visually distinguishable avatars. Our virtual bar itself was purchased from a vendor in SL and populated with objects commonly found in bar settings. Posters, dart boards, sofas, tables, lights, a dance stage, and stock of drinks are all present in the experiment's virtual environment. As seen in Figure 1, the bar consists of areas appropriate for the waiter/waitress (behind the bar), places for the regulars and their associates to group (tables, bar stools, sofa), out of the way locations for the wall flowers (places to sit in the corners and pools to lean on), and an area for showing off and dancing (the dance floor).

Scripted Bots

For our studies, we wished to compare the performance of human confederates and SCR Avatars with a traditional scripted bot found in the background environment of many modern game engines and virtual worlds. The publicly available bots in the SL market are typically primitive vending machines for drinks or dances that wait for SL users to choose from a scripted menu of items, and may print out a static piece of text for new visitors, or repeated pieces of text for existing guests. Because these primitive existing bots did not have any concept of sustained and active interaction with people around them, we implemented our own bots to represent a typical scripted bot from other game environments.

For the observational study, we implemented two bot roles, the Wallflower and the Waiter/Waitress, and then added a Generic role to our full study. The traditional scripted bots, unlike the ABL-based SCR Avatars, were functionally simple and did not require the language and infrastructure features of ABL. They were constructed based on pre-authored finite state machines, containing list of behavior states and associated ordered actions (such as playing one of three gestures, giving a drink to a patron, walking to a new location, or dancing) that changed and repeated over time, with random variations in behavior on each step and random time intervals between behaviors in order to avoid obvious repetitive behavior patterns..

Observational Study

We began with an Observational Study, a non-intrusive evaluation which placed 7 participants with confederates and traditional scripted bots in our scenario. These participants were given the task list described a previous section, and given unlimited time to complete these tasks. In order to account for gender biases affecting the perceived personality and interactions of the various patrons of the bar, the gender of each avatar, as evident by their physical appearance and clothing, was set based on the gender of the participant, either 'same' gender as the subject, or 'opposite' gender from the participant. This study included 8 bar patrons, with 3 human-controlled Regular avatars, 1 human-controlled Spring Breaker avatar, a human and computer controlled Waiter/Waitress, and a human and computer controlled Wallflower.

Each of our confederates were asked to freely role-play an assigned social role in our virtual bar using all available interactions except chat, while our computer-controlled avatars used a fixed set of sequential actions based on these roles. The protocol followed by this study consisted of the following procedure: (1) Participant reads and signs informed consent document, (2) SL tutorial and overview of the participant task list, (3) Setup virtual world with confederates and computer-controlled avatars, and enable data logging, (4) Perform experiment in SL, (5) Debriefing interview, and (6) Survey.

Using video recording and automatic data logging using in-game logging scripts, we observed and analyzed the interactions amongst all avatars, both human and computer-controlled. After running our experiments, we viewed the captured video and noted actions or style of actions that were significantly different toward bots than toward human avatars, and analyzed our data logs using various metrics gathered from previous virtual world interaction research. These metrics, discussed in detail in our results section, include interpersonal distances and space categories, gaze angles and sums, and measures of presence and co-presence. This initial study allowed us to confirm that there were measurable, significant differences between human-human and human-computer avatar interaction in our scenario, motivating our further experiments. In Figure 2, we show a spectrum of Avatar Realism that we hypothesized would exist, ranging from the realism of a traditional scripted bot to that of fully human-controlled avatar. In our research, we expected that our metrics would place an agent with effective Supporting Character Realism in between these two ends of the spectrum.

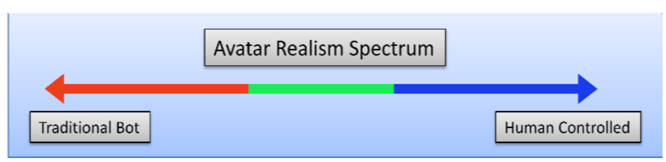

Figure 2. The spectrum of Avatar Realism. We expect the metrics of an agent with effective SCR to be placed between the realism of a traditional bot and that of a human controlled avatar.

Our Observational Study, in addition to establishing a concrete way to measure the differences in interaction between computer and human controlled avatars, provided a large catalog of recorded behavior of humans enacting social roles in our scenario. We used this information to design computer-controlled avatars designed to exhibit effective SCR - our SCR avatars. In our study, we placed these AI-driven avatars alongside confederates and bots in the same social scenario as our Observational Study, in order to evaluate their realism and place them somewhere in the spectrum shown in Figure 2.

SCR Avatar Technology

Our prototype SCR Avatars were designed directly from the observed and measured behaviors of the human confederates in our observational study. We implemented these agents in a reactive planning language called *A Behavior Language*, or *ABL* [17]. This language and its associated infrastructure is designed to support virtual agents that have a large number of sensors that read information from the world (in our case, SL), and react to them quickly, and often simultaneously, with a series of actions.

In our scenario, the names, social roles, locations, movements, and gestures of each avatar in our bar were taken as input and fed into a set of behaviors acting in parallel, constantly analyzing, storing and reacting to this data with actions that included all of the possible types of interactions available to any human in our scenario. We built discrete sets of behaviors that matched the various behavior modes we observed in our Observational study, such as: getting a drink, dancing with a group, initiating and responding to various types of gestures, and moving around the bar.

In addition, in order to allow our avatars to blend in well with human-controlled avatars, we designed our SCR Avatars to exhibit the quirks of human control of a Second Life avatar, including imprecise and occasional body rotation and movement, delays in gesture responses and recognition, and unrealistic head movement to match the effect of a user moving his or her mouse (which controls the head of the avatar in the Second Life Viewer) to the gesture list or button before activating it.

Full Study

As a trial prior to our full study, we conducted a pilot study, consisting of 5 participants (4 male, 1 female) to assess how the procedure and tasking from our previous observational study needed to be tweaked given the introduction of SCR avatars to the virtual bar. We did not conduct full statistical analysis of the pilot study; the main goal was to iron out procedures and get feedback from participants. As planned, we added our newly implemented SCR avatar personalities for the pilot: a Regular, a Spring Breaker, and a Waiter/Waitress, and given the participant feedback and our analysis in the Observational Study, we added the Generic bot role and removed the human Wallflower role, for a total of 11 confederate avatars comprised of five human confederate avatars, three SCR avatars and three bots. Pilot testing showed that it was confusing to have several waiters in the small space, and that overall there were too many other avatars to 'get to know' in a short period of time. As a result , we further refined the roles for our full study by removing two waiter/waitress avatars.

Our full study, with the additional SCR avatar roles and feedback from both our observational and pilot studies, consisted of: 3 human confederate controlled avatars (2 Regulars and 1 Spring Breaker), 3 SCR avatars (1 Regular 1 Spring Breaker, and 1 Waiter/Waitress), and 2 Bot controlled avatars (1 Wallflower and 1 Generic). We had a total of 24 participants (4 females and 20 males) over a span of 3 sessions. Additional changes to the full study included a simplified and compacted participant task list, overhead video capture view of the scenario to supplement participant view capture, and the addition of a time limit as an independent variable. From our previous analyses, we had identified that time in the bar could have an impact on a subject's interactions with other avatars, and so we set explicit time limits for the amount of time participants had in the virtual bar - half of the participants had a 12 minute time limit to complete the task list, while the other half had 24 minutes.

Metric	Description
Gaze Angle	The angle an avatar faces relative to the direction towards a target. It ranges from 0 (facing target) to 180 (turned directly away from target) [22]
Gaze Sum	Sum of the gaze angles between two avatars; it ranges from 0 (two avatars facing each other) to 360 (two avatars looking completely away from each other). [22]
Interpersonal Distance (IPD)	Pair-wise distance between avatars
Space	Categorized as public, social, personal, or intimate [11]
Gesture Target	The average number of gestures targeted at an avatar
Co-presence	The extent to which participants reported behaving and responding as if the avatars were real [18]

Table 1: Candidate metrics based on observational study results

Measures

A variety of social interaction dimensions exist which can be used to evaluate the interactions in our experiments. For example, Blascovich et. al.'s threshold model of social influence [4] considers social presence and behavioral realism when considering interactions between avatars in virtual world settings. Under this framework, categories of observable behaviors can be used to predict the level of social influence a person might be experiencing. For example, prior work by Bailenson [2] has shown that interpersonal distance is one such reliable indicator: people move closer to avatars that have a lower level of realism than those with a higher level of realism. Our observational study results presented a set of viable candidate metrics which showed measurable differences between the human and bot controlled avatars, shown in Table 1. We used these metrics as a potential indicator of realism in our full study.

RESULTS AND ANALYSIS

In Table 2, we summarize the results of the within group analysis of variance (ANOVA) for the following measures:

- Gaze Angle: from the participant to the various avatars
- Gaze Sum: Two avatars facing towards each other
- Interpersonal Distance: pair wise distance between avatars
- Gesture Target Towards Avatar
- Co-presence

Analysis was conducted for the main effects of:

- Avatar Type (3) – Human, SCR avatar, and Bot
- Avatar Role (8) – looking at the 8 avatars individually

Measure	Effect	df	F	Sig.	Power
Gaze Angle	Type	1.61	502.62	.000	1.00
	Role	4.31	277.23	.000	1.00
Gaze Sum	Type	1.84	1009.04	.000	1.00
	Role	6.05	728.97	.000	1.00
IPD	Type	1.58	1968.22	.000	1.00
	Role	4.06	2410.23	.000	1.00
Gesture Target	Type	1.77	10.71	.000	.97
	Role	3.68	7.51	.000	.99
Co-Presence	Type	1.97	2.76	.075	.51
	Role	4.74	10.54	.000	1.00

Table 2. Statistics summaries

The *F* tests that are reported for the within group effects include the Greenhouse-Geisser correction when necessary to protect against possible violation of the sphericity assumption. All reported significance values are at $p < .001$ with an observed power > .95, unless otherwise stated.

Follow-up within subject contrasts were conducted using pairwise comparisons. Significance for pairwise comparisons are reported at a p <.001 level of significance.

A 4 (space) x 3 (avatar type) repeated measures analysis was conducted for the Space metric to look at possible interaction effects of social space and avatar type. These results are not included in Table 2.

Gaze Angle

From the captured data, we calculated the gaze angles (derived from previous work by Yee et al) between the participant avatar and all other avatars [23]. A lower gaze angle means that the participant was more often looking more directly at an avatar. Figure 3 shows the results distributed by avatar type.

This figure illustrates that participants maintained the smallest gaze angle with human avatars, followed by SCR avatars, and the largest with bots. The main effect of avatar type on gaze angle was significant (see Table 2). In addition, follow-up within subject contrasts showed that all pairwise comparisons were significant. We can see here that SCR avatars performed closer to human controlled avatars than traditional bots maintaining a more direct gaze at our participant, similar to our human confederates in the scenario. This move towards the human end of the spectrum benefits our target of SCR, since we expect these prototype avatars to that of human controlled avatars.

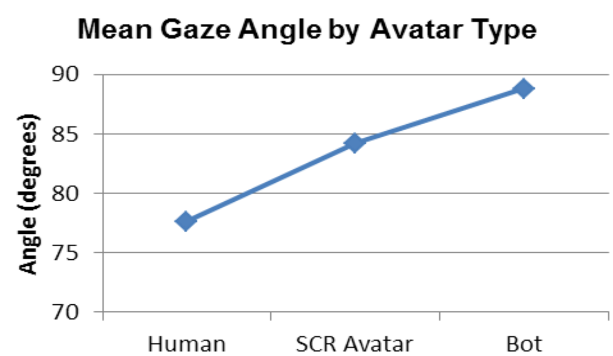

Figure 3. Gaze Angle means from the Participant to Avatars by Type

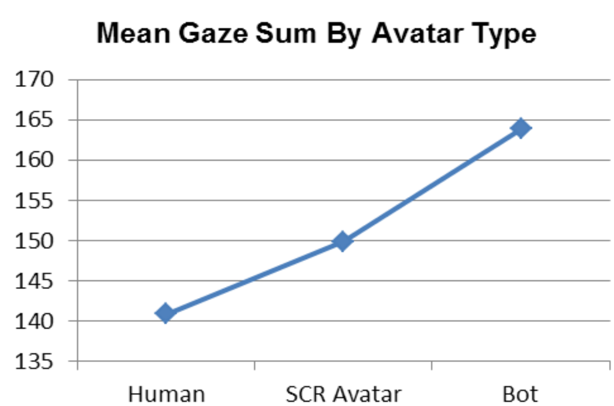

Figure 4. Gaze Sum means between the Participant and Avatars by Type

Gaze Sum

The Gaze sum is a combination of two gaze angles between two avatars, as proposed by Yee et al [23]. As mentioned in a previous section, this measurement quantifies how much two avatars are looking at each other, with zero degrees representing two avatars looking directly at each other, and 360 degrees corresponding to two avatars looking directly away from each other. In this analysis, we measured the mean gaze sum of each avatar through runs of our studies by combining the gaze angles between each avatar type and the participant towards that specific avatar, and grouped these sums by type.

Figure 4 illustrates a significant effect for avatar type on gaze sum. Follow-up within subject contrasts showed that all pairwise comparisons were significant. Like with our gaze angle measurements, the mean gaze sum similarly shows that our SCR avatars place between the performance of a human and bot, approaching the target of human-level behavior.

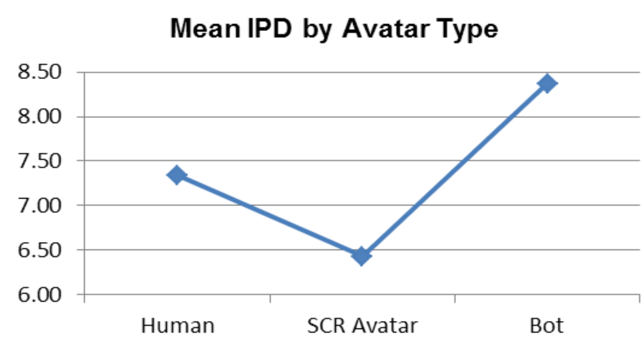

Figure 5a. Interpersonal Distance means by Avatar Type

Interpersonal Distance (IPD)

Interpersonal distance (or IPD) from the participant was calculated (in meters) from the positional data that was logged for all avatars. Both avatar type and avatar role had a significant effect for avatar type on interpersonal distance. Figure 5a shows that our SCR avatars tended to maintain a smaller distance to the participant compared to Bots, and Figure 5b shows that our Spring Breaker almost always was closer to the participant than any Human or Bot, due to particularly aggressive social behaviors that constantly aimed to dance with all of the patrons of our virtual bar. In future work, in order to blend in with the human confederate avatars, this particularly outstanding behavior would be modified to more closely match the distances achieved by the human Spring Breaker.

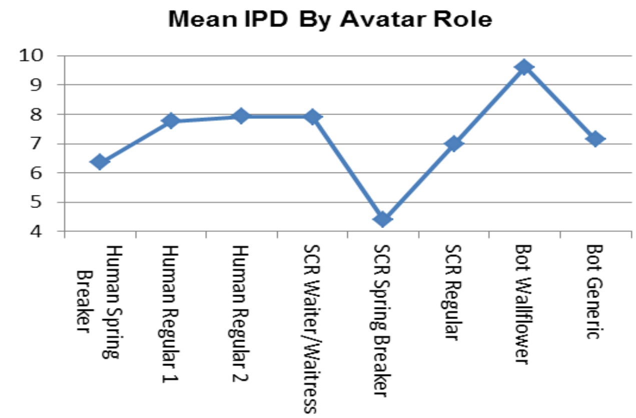

Figure 5b. Interpersonal Distance means by Avatar Role

Space

Space was calculated using the personal reaction bubbles developed from Hall [11]. The interpersonal distance measures were then grouped into four bins: Intimate space (< 0.45 m), Personal Space (< 1.2 m), Social Space (< 3.6m), and Public Space (> 3.6 meters).

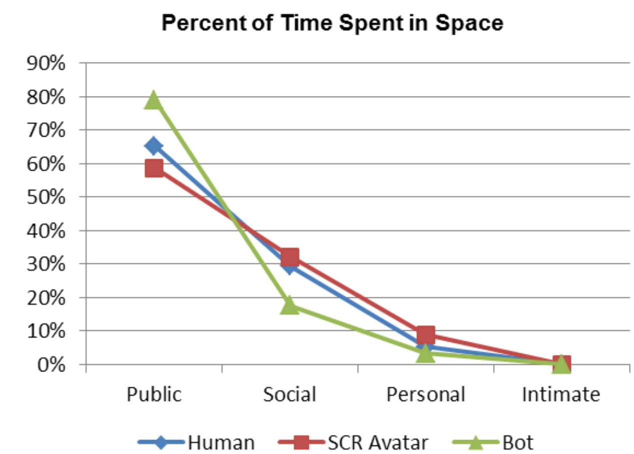

Figure 6. Percent of time spent in various space categories by Avatar Type

A 4x3 (space categories by avatar type) repeated measures analysis demonstrated a significant interaction effect for avatar type on space (see Table 2. Participants spent most of their time in public space with all avatars, but this was highest for Bots. Figure 6 shows that our SCR avatars tended to stay at the same distances from our participants for the same percentage of time as our human confederates.

Gesture Target

Gesture targets towards from the participant to other avatars were calculated by finding the nearest avatar facing the participant when the participant was using a gesture. The avatar that was facing the participant and closest to him or her was counted as the target for that particular gesture.

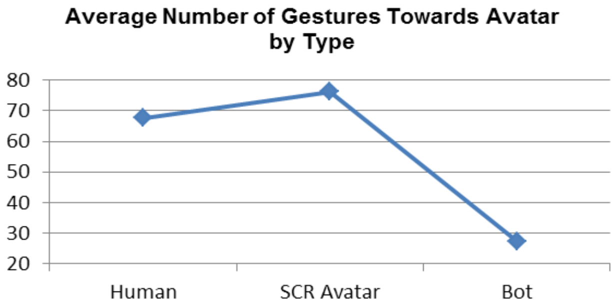

Figure 7a. Average Number of Towards Avatars by Types

Both avatar type and avatar role had a significant effect on the average number of gestures used by participants (Table 2). Follow-up within subject contrasts showed that there was a significant difference between human and bot, and SCR avatar and bot, but the difference between humans and SCR avatars was not significant (Figure 7a). A repeated measures analysis demonstrated a significant effect for avatar role on gestures (Table 2). Participants gestured significantly more to the SCR Spring Breaker than any other avatar (Figure 7b) - a result of this avatar's aggressive social behaviors towards the participant and other avatars in the bar.

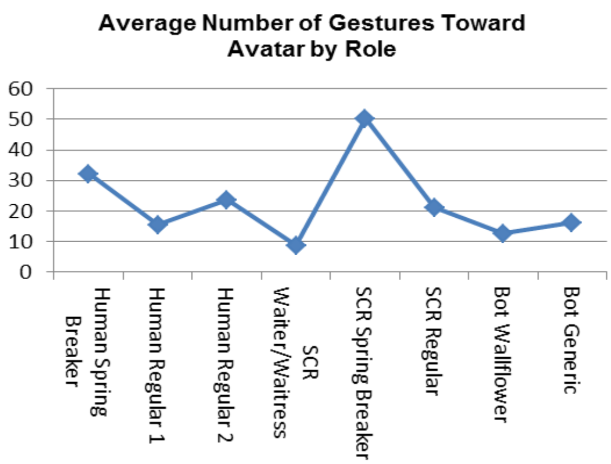

Figure 7b. Average Number of Gestures Towards Avatars by Role

Co-Presence

Co-presence refers to the participants' sense of being with another person in the virtual bar. The co-presence questionnaire was adapted from the Slater Co-Presence Questionnaire [18]. Participants used a 7-point numerical scale (1. not at all to 7. a great deal) to respond to seventeen items (such as "I had a sense of being with the other person..." or "The experience seems to me more like interacting with a person…"). Responses to the items were used to compute co-presence mean, which was the average rating across all of the items.

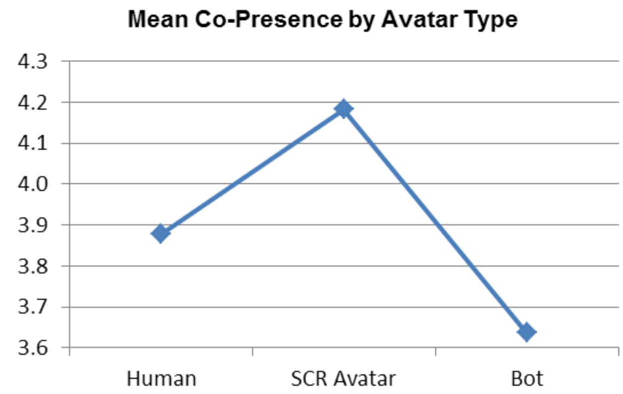

Figure 8a. Co-presence means by Avatar Type

Although the effect of avatar type on co-presence mean was not significant, co-presence was highest for the SCR avatars (M=4.18), followed by human (M=3.88), and lowest for bots (M=3.64) (Figure 8a). Our analysis also shows a significant effect of avatar role on co-presence (Table 2). While not all of our prototype avatars scores significantly better than other roles, the sense of co-presence was significantly higher for the SCR Spring Breaker (Figure 8b), showing great potential for the for future development of other roles that learn from the lessons of our Spring Breaker's design.

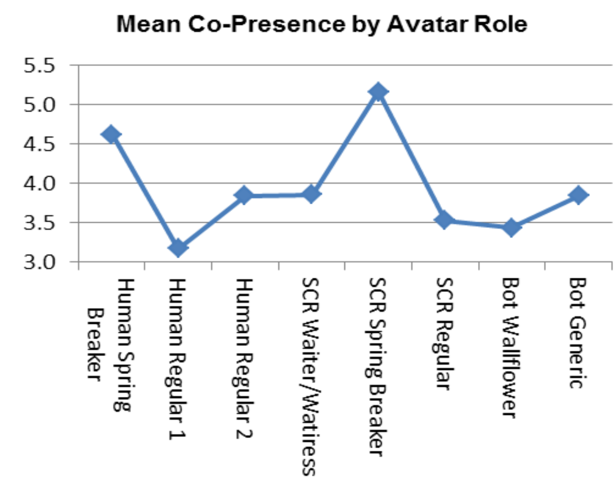

Figure 8b. Co-presence means by Avatar Role

Debriefing and Comments

During the participant debriefing sessions, several of the participants indicated that they would have really like to be able to use chat in the virtual bar in order to complete the assigned tasks. Participants also felt that chat would have made their experience much more realistic. Overall, participants found the experience "cool", "fun", and "interesting". However, they felt it was hard to understand the reactions of the other avatars. They also had trouble knowing whether or not other avatars "received/acknowledged" gestures intended for them.

Participants thought the other avatars were realistic with distinguishable personalities. However, many felt that the virtual bar was not crowded enough, and would have liked to have a live band, more people, etc.

The following are sample participant comments from our debriefing sessions (note that Pat, Harper, and Alec were SCR avatars in our full study):

- "…Pat, Ashe and Terry were a tight knit group. Quinn was that awkward newbie. Devin, was the one used to the world and was just having fun doing her own thing. Harper was fun, but a little outrageous. Terry...that smug b****. Parker, that goth non-conformist. Pat...Pat was Pat. He had no characteristics, but then again he was in a tight knit group"
- ".. one point I was wondering if someone was actually playing the people. Their responses to my actions and movements made the game pretty fun. ... Ashe was the social one, but was distant to the new guy …Terry was the same, and was in the same group as her. Devin was the dancer, she just danced, usually by herself. Alec was the guy who you had to get to know to have him like you. Quinn was the person who just right off the bat gave you a drink, and Harper was the person who immediately started dancing with you. Add in Pat who refuses to dance with other people, and Parker who just refused to even look at me, and you have an entire cast of people who you can find in high school. However, they were believable characters, and I did come up with all this after only playing a few minutes, even if I'm off I still have a firm idea of what kind of character they are..."

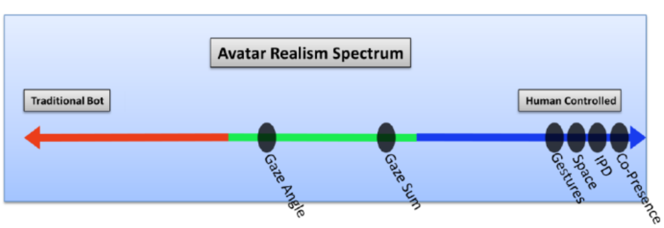

Figure 9. An illustration of relative performance our SCR avatars for each metric

DISCUSSION

In the observational study, we were able to establish both objective (gaze angle and sum, IPD, gesture target, and space categorization) and subjective (co-presence) baseline metrics as strong indicators of agency in comparing human controlled and bot controlled avatars. The goal of the full study was to evaluate our prototype SCR avatar with respect to those established candidate metrics, and see where these metrics fall on the "Avatar Realism Spectrum" for the SCR avatars (Figure 2).

Our hope was that our SCR avatars would score closer to human avatars than traditional bot avatars. As you can see from our results, we did just this: our prototype avatars *scored closer to humans than to bots with statistical significance* for all analyses in our metrics but co-presence, and our SCR Spring Breaker showed a significantly higher co-presence for his or her role than any other character in our scenario. Figure 9 illustrates relative performance for each metric using an ad hoc summary comparison.

Duration Effects

In general, our longer 24 minute study runs showed significant effects on gaze angles, gaze sums and interpersonal distances with our SCR avatars, with higher values on all of these measures. This effect on duration likely shows that as participants spent longer periods in our virtual bar, they spent more time with our avatars and lost interest with these interactions, stressing the capabilities of our SCR-focused behaviors. Extended direct interaction time, even when it does not include language interaction, is a condition that begins to move out of the realm of "supporting character." Future work would involve examining where and how the behaviors of our SCR avatars cause participants to start losing interest (perhaps suspecting that they are computer controlled). These longer runs showed no significant effects on gesture target measures, and the lack of significance for our co-presence measures for all participants did not warrant analysis based on duration.

Gender Effects

In general, females in our study tended to treat our SCR avatars more like human controlled avatars, with lower gaze angles and sums, closer interpersonal distances, and a higher sense of co-presence. However, since we did not have enough female participants to demonstrate significant effects, we leave a full analysis of these gender effects to future work.

LIMITATIONS AND FUTURE WORK

Our work here shows that, given our baseline metrics, SCR avatars can perform comparably to human-controlled avatars in performing supporting character roles, and almost always significantly better than traditional scripted bots. Future work should examine the effects of gender on the perception of co-presence and realism that were hinted at in this work. The behavioral repertoire of SCR avatars should be expanded to support a wider palette of physical interactions. While SL provides a large number of gestures and the ability to add additional user-provided animations, it contains lower fidelity and primitive animation and interaction functionality compared to many contemporary game engines. The addition of full facial expressions, higher resolution animation skeletons and models, accurate and realistic physical collision, and more nuanced animations with smooth transitions and blending would allow our agent to explore a much larger pallet of social interaction, and allow our participants to more easily perform tasks by accurately gauging avatar responses and subtle emotional states. These interactions could include subtle changes in posture and facial expression, direct touches or hugs, and layered emotional responses that can be expressed differently based on their location in the virtual body. Finally, future work should explore the possibility for limited language capability of an agent targeting SCR. Establishing metrics that evaluate the linguistic skills of a character for the purposes of SCR will be critical in this future work.

CONCLUSION

In this paper, we have established SCR both conceptually and experimentally. We first proposed the concept of SCR and distinguished it from believability and Turing Test realism, and created an experimental scenario that allows human participants to interact with mixed groups of humans, bots and SCR avatars. By analyzing the data from the first of these experiments, we then established a set of candidate metrics to measure the interaction amongst humans and computer controlled avatars. Finally, we showed that our prototype SCR avatars significantly outperformed traditional scripted bots and approached the interaction realism of the human confederates playing supporting characters. Ultimately, this work provides both a method of defining and measuring SCR, and a demonstration that it is technically possible to approach SCR with contemporary technology.

REFERENCES

1. Bates, J. The role of emotion in believable agents. *Communications of the ACM 37*, 7 (1994), 122-125.
2. Bailenson, J.N., Beall, A.C., Loomis, J., Blascovich, J., and Turk, M. Transformed Social Interaction: Decoupling Representation from Behavior and Form in Collaborative Virtual Environments. *Presence: Teleoperators and Virtual Environments 13*, 4 (2004), 428-441.
3. Bell, D. Learning from Second Life. *British Journal of Educational Technology 40*, 3 (2009), 515-525.
4. Blascovich, J., Loomis, J., Beall, A.C., Swinth, K.R., Hoyt, C.L., and Bailenson, J.N. Immersive Virtual Environment Technology as a Methodological Tool for Social Psychology. *Psychological Inquiry 13*, 2 (2002), 103-124.
5. Brookey, R.A. and Cannon, K. Sex Lives in Second Life. *Critical Studies in Media Communication 26*, 2 (2009), 145-164.
6. Buono, P., Cortese, T., Lionetti, F., Minoia, M., and Simeone, A.L. A Simulation of a Fire Accident in Second Life. *Proceedings of the 11th Conference on Presence*, CLEUP, Padova, Italy (2008), 183-190.
7. Burke, M., Kraut, R., and Marlow, C. Social capital on facebook. *Proceedings of CHI ’11*, ACM Press (2011), 571.
8. Cheung, G. and Huang, J. Starcraft from the stands: understanding the game spectator. *Proceedings of CHI ’11*, ACM Press (2011), 763.
9. Coleridge, S.T. Biographia Literaria (1817). *Search*, (2006), 4.
10. Friedman, D., Steed, A., and Slater, M. Spatial Social Behavior in Second Life. *Intelligent virtual agents 4722*, (2007), 252-263.
11. Hall, E.T. *The Hidden Dimension*. Doubleday, 1966.
12. Harris, H., Bailenson, J.N., Nielsen, A., and Yee, N. The Evolution of Social Behavior over Time in Second Life. *Presence Teleoperators Virtual Environments 18*, 6 (2009), 434 - 448.
13. Huvila, I., Holmberg, K., Ek, S., and Widén-Wulff, G. Social capital in Second Life. *Online Information Review 34*, 2 (2010), 295-316.
14. Inman, C., Wright, V.H., and Hartman, J.A. Use of Second Life in K-12 and Higher Education : A Review of Research. *Review Literature And Arts Of The Americas 9*, 1 (2010), 44-63.
15. Lampe, C.A.C., Ellison, N., and Steinfield, C. A familiar face(book): profile elements as signals in an online social network. *Proceedings of CHI ’07*, ACM Press (2007), 435.
16. Laurel, B.K. Toward The Design Of A Computer-Based Interactive Fantasy System. *Drama Department*, 1986, 312 ST - Toward the Design of a Computer-Based In.
17. Mateas, M. and Stern, A. A Behavior Language: Joint action and behavioral idioms. *Life-like Characters. Tools, Affective Functions and Applications 194*, (2004), 1-28.
18. Mortensen, J., Vinayagamoorthy, V., Slater, M., Steed, A., Lok, B., and Whitton, M.C. Collaboration in tele-immersive environments. (2002), 93-101.
19. de Pascale, M., Mulatto, S., and Prattichizzo, D. Bringing haptics to Second Life. *Proceedings of the 2008 Ambi-Sys Workshop on Haptic User Interfaces in Ambient Media Systems*, (2008), 6.
20. Taylor, T.L. *Play Between Worlds: Exploring Online Game Culture*. MIT Press, 2006.
21. Turing, A.M. Computing Machinery and Intelligence. *Mind 59*, 236 (1950), 433-460.
22. Weitnauer, E., Thomas, N.M., Rabe, F., and Kopp, S. Intelligent Agents Living in Social Virtual Environments – Bringing Max Into Second Life. *Germania*, (2008), 552-553.

23. Yee, N., Bailenson, J.N., Urbanek, M., Chang, F., and Merget, D. The unbearable likeness of being digital: the persistence of nonverbal social norms in online virtual environments. *Cyberpsychology and Behavior 10*, 1 (2007), 115-121.

24. Yee, N. and Bailenson, J.N. A method for longitudinal data collection in Second Life. *Presence Teleoperators Virtual Environments 17*, 6 (2008), 1-6.

25. Yee, N., Harris, H., Jabon, M., and Bailenson, J.N. The Expression of Personality in Virtual Worlds. *Social Psychological and Personality Science 2*, 1 (2010), 5-12

26. Yee, N., Ducheneaut, N., Yao, M., and Nelson, L. Do Men Heal More When in Drag ? Conflicting Identity Cues Between User and Avatar. *Proteus*, (2011), 773-776.

27. Zebrowitz, L.A. The Affordances of Immersive Virtual Environment Technology for Studying Social Affordances. *Psychological Inquiry 13*, 2 (2009), 143-145