Three Avatars Walk into a Bar: Defining and 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. 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 social interaction experiment set in a virtual bar amidst a variety of human and computer controlled patrons. Finally, we employ these metrics to measure the SCR performance of traditional scripted bots and prototype AI-driven agents. 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.

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

General Terms
Design, Experimentation, Measurement

Keywords
Avatars, Realism, Virtual worlds, Artificial Intelligence.

1. 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 [6, 13]; virtual story worlds for entertainment and game play purposes, such as massively-multiplayer online roleplaying games [16], competitive multiplayer games that include thousands of audience members [7] and education and training [12]. Experiential training focused on social skills and human interaction, such as 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 does not 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.

While there have been many virtual world studies looking at human-to-human interaction (see Background) very few studies have looked at mixed human and NPC interaction. 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 be refined 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, NPCs 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. 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 do not immediately highlight their representational nature. In contrast, *Turing Test realism* assumes the adversarial conditions of the Turing Test [17], 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. Support for skeptical probing is not required for most game applications, and, as it likely requires a complete AI model of a human being, is not achievable using current technology.

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 is not intended to engage in long-term complex behavior with the player, but rather to provide a sense of realistic human activity around the

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.
FDG'12, May 29 – June 1, 2012, Raleigh, NC, USA.
Copyright 2012 ACM X-XXXXX-XXX-X/XX/XX12…$10.00.

player. In a mixed environment of human avatars and SCR avatars, in the context of background activity and lightweight interaction with human-controlled avatars, an SCR avatar blends in seamlessly. In this paper we develop metrics for measuring SCR and demonstrate the technology feasibility of achieving it.

To explore SCR, AI-controlled prototype avatars were developed and tested 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 SL scripting techniques. We first established baseline metrics used 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 tasks of social interaction. If successful, 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 majority of SCR metrics.

2. 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.

SL and other virtual worlds have been used as a methodological tool to study human interactions with both other human and AI controlled avatars [3]. Studies in the social behavior of players, such as accrual of social capital [11], gender and sex practices in virtual worlds [4], social affordances of players [23], decoupling of non-human behavior between players [5], and effects of gender differences between a human and their avatar [22] 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 [8]. 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 of players changes over time and expressions of personality in virtual worlds were analyzed in a longitudinal study conducted in SL [10, 21]. A framework to gather avatar-related information from SL over extended periods of time [20] was used in conjunction with qualitative measures to perform both analyses. We gather similar avatar data, though details of our collected metrics and analysis methods differ; additionally, we gather such data for bots and SCR avatars as well as humans.

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

Weitnauer [18] 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 *libOpenMetaverse*[2] framework to connect our AI system and data reporting framework to SL. Weitnauer's work is a technical proof of concept, and thus was never evaluated with respect to human interaction.

3. 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.

3.1 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, and interacts with the following roles:

Regular: The veteran of the nightclub/bar. The regular always is among others of his or her kind and knows 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

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

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.

3.2 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.

3.3 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 for the creation of a plausible, furnished and decorated testing environment filled with visually distinguishable avatars. As seen in Figure 1, the virtual bar testing environment consists of areas appropriate for scenario and character roles.

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

3.4 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.

3.5 Observational Study

We began with an observational study, a non-intrusive evaluation in which 7 participants interacted with confederates and traditional scripted bots in our scenario. These participants were given a task list and were permitted unlimited time to complete their tasks. 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. This study included 8 bar patrons, with 3 human-controlled Regular avatars, 1 human-controlled Spring Breaker avatar, a human and bot controlled Waiter/Waitress, and a human and bot controlled Wallflower.

The confederates were asked to role-play an assigned role in our virtual bar using all available interactions except chat, while our bot-controlled avatars used a fixed set of sequential actions based on these roles.

Using video recording and automatic data logging we observed and analyzed the interactions amongst all avatars, both human and computer-controlled. The captured video was reviewed and actions or style of actions that were significantly different toward bots than toward human avatars were noted. The data logs were analyzed 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. We hypothesized that there is a spectrum of Avatar Realism, ranging from the low realism of a traditional scripted bot to that of a fully human-controlled avatar, and that effective SCR lies in between these two ends of the spectrum.

Our observational study also provided a large catalog of recorded behavior of humans enacting social roles in our scenario. We used this information to create computer-controlled avatars designed to exhibit effective SCR – our prototype SCR avatars. In our full study, we placed these AI-driven SCR avatars alongside confederates and bots to evaluate their realism.

3.6 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 avatars in a reactive planning system called *A Behavior Language*, or *ABL* [14]. The ABL 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 each of 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 in the SL viewer interface in order to activate a chosen gesture. It is important to note that the behaviors of our SCR avatars could not feasibly be implemented using traditional bot scripting approaches. As this paper focuses on our study methods and results, we do not provide the specific technical details of our SCR avatars in this paper.

Table 1. Candidate metrics based on observational study results

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) [18]
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). [18]
Interpersonal Distance (IPD)	Pair-wise distance between avatars
Space	Categorized as public, social, personal, or intimate [9]
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 [15]

3.7 Full Study

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. The full study included a simplified 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.

3.8 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 [3] 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.

4. RESULTS AND ANALYSIS

In Table 2, we summarize the results of the within group analysis of variance (ANOVA) for the measures referred to in the previous section: Gaze Angle, Gaze Sum, Interpersonal Distance (IPD), Gesture Target, and 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

Table 2. Statistics summaries

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

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.

4.1 Gaze Angle and Gaze Sum

From the captured data, we calculated the gaze angles and gaze sums (derived from previous work by Yee et al) between the participant avatar and all other avatars [19]. A lower gaze angle means that the participant was more often looking more directly at some other avatar, and a lower gaze sum indicates that both the participant and some other avatar were more directly looking at each other.

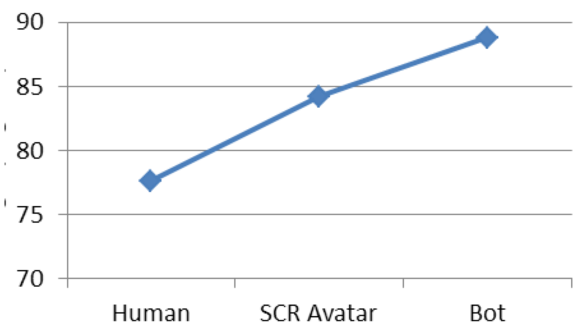

Figure 2. Gaze Angle (in degrees) means from the Participant to Avatars by Type

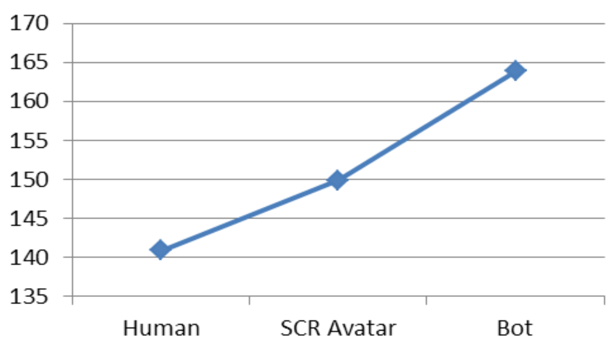

Figure 3. Gaze Sum means between the Participant and Avatars by Type

Figure 2 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 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.

4.2 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 4a shows that our SCR avatars tended to maintain a smaller distance to the participant compared to Bots, and Figure 4b 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.

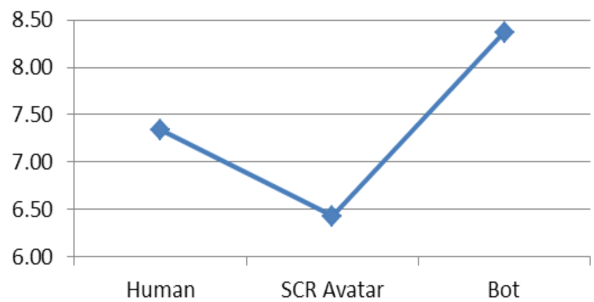

Figure 4a. Mean Interpersonal Distance (in meters) by Avatar Type

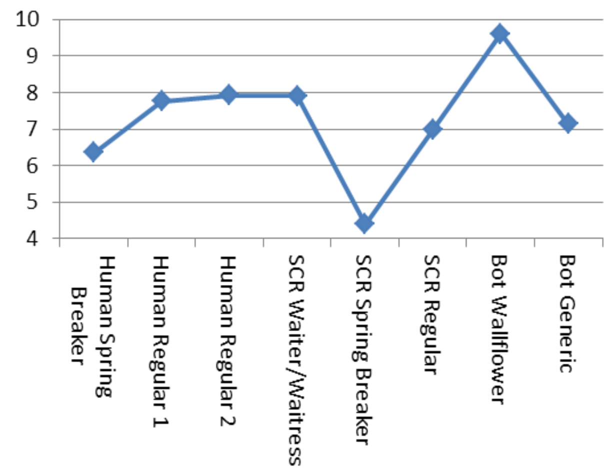

Figure 4b. Mean Interpersonal Distance (in meters) by Avatar Role

4.3 Space

Space was calculated using the personal reaction bubbles developed from Hall [9]. 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).

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 5 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.

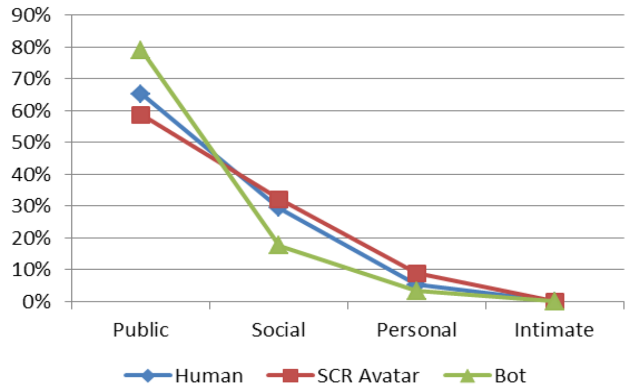

Figure 5. Percent of time spent in various space categories by Avatar Type

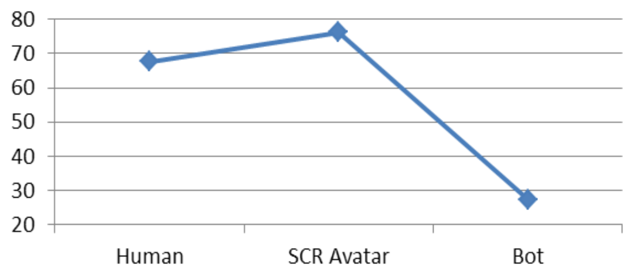

Figure 6a. Average Number of Gestures Towards Avatars by Types

4.4 Gesture Target

Gesture targets from the participant to other avatars were calculated by selecting the nearest avatar that the participant is facing while using a gesture. Figure 6a shows the average number of gesture targets for all participants, separated by avatar type, while Figure 6b separates these targets by role.

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 6a). 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 6b) - a result of this avatar's aggressive social behaviors towards the participant and other avatars in the bar.

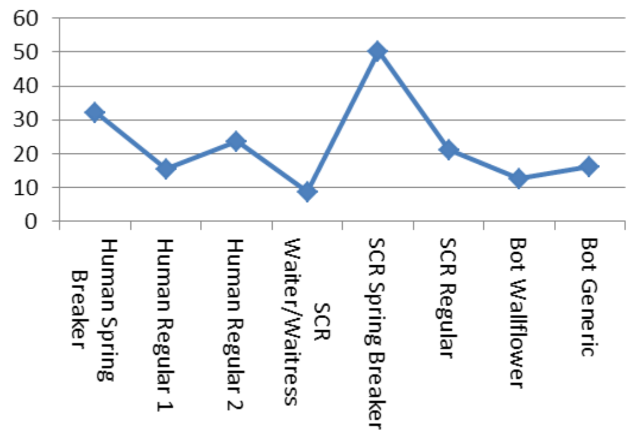

Figure 6b. Average Number of Gestures Towards Avatars by Role

4.5 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 [15]. 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.

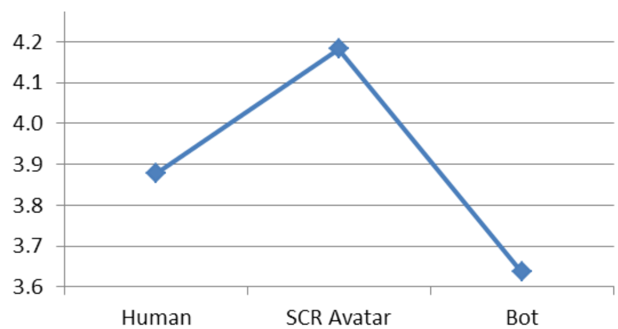

Figure 7a. Co-presence means by Avatar Type

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 7a). 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 7b), showing great potential for the for future development of other roles that learn from the lessons of our Spring Breaker's design.

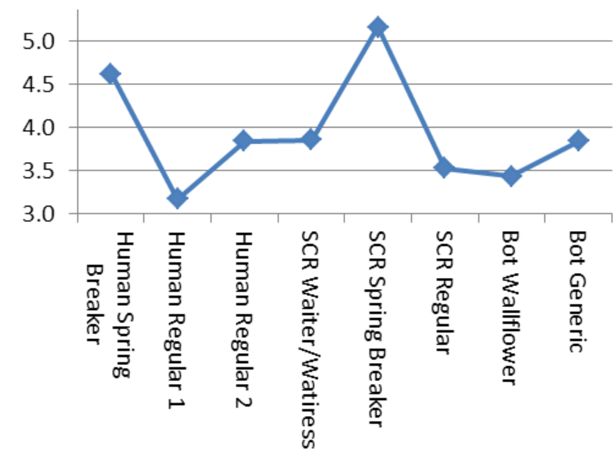

Figure 7b. Co-presence means by Avatar Role

5. 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.

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 8 illustrates relative performance for each metric using an ad hoc summary comparison.

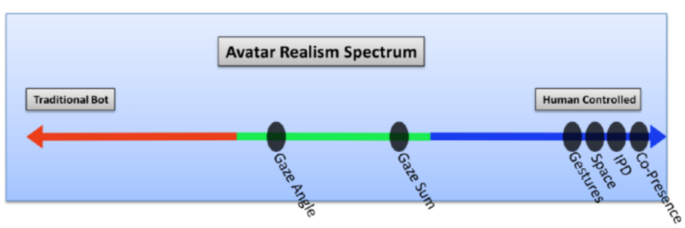

Figure 8. An illustration of relative performance our SCR avatars for each metric

5.1 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 of time in our virtual bar, they similarly 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 extend beyond the realm of "supporting character" tasks. 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).

5.2 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.

6. 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. In a future study we would like to verify these results with an across-subjects study in which individual participants are exposed to either a solely bot, solely human confederate, or solely SCR avatar environment. A challenge for performing this study is that it is not feasible to replicate the richness of behavior of our ABL-based SCR avatars with traditional bot scripting technology, and thus this study will face the question of what constitutes "best effort" in implementing the bot condition. Additionally, further evaluations should fully examine the potentially interesting effects of gender on the perception of co-presence and realism that were hinted at in this work.

In addition to revisions to our evaluation method, future work should aim to expand the behavioral repertoire of SCR avatars through support for 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, complex inter-avatar physical interactions, higher resolution animation skeletons and models, and more nuanced motions with smooth transitions and blending would allow our SCR avatars to explore a much broader spectrum of social interaction, and allow our participants to more easily perform tasks by accurately gauging avatar responses and subtle emotional states. 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.

7. 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.

8. ACKNOWLEDGMENTS

This research is sponsored by the Air Force Research Laboratory.

9. 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] 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.

[4] Brookey, R.A. and Cannon, K. Sex Lives in Second Life. *Critical Studies in Media Communication 26*, 2 (2009), 145-164.

[5] 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.

[6] Burke, M., Kraut, R., and Marlow, C. Social capital on facebook. *Proceedings of CHI '11*, ACM Press (2011), 571.

[7] Cheung, G. and Huang, J. Starcraft from the stands: understanding the game spectator. *Proceedings of CHI '11*, ACM Press (2011), 763.

[8] Friedman, D., Steed, A., and Slater, M. Spatial Social Behavior in Second Life. *Intelligent virtual agents 4722*, (2007), 252-263.

[9] Hall, E.T. *The Hidden Dimension*. Doubleday, 1966.

[10] 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.

[11] Huvila, I., Holmberg, K., Ek, S., and Widén-Wulff, G. Social capital in Second Life. *Online Information Review 34*, 2 (2010), 295-316.

[12] 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.

[13] 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.

[14] 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.

[15] Mortensen, J., Vinayagamoorthy, V., Slater, M., Steed, A., Lok, B., and Whitton, M.C. Collaboration in tele-immersive environments. (2002), 93-101.

[16] Taylor, T.L. *Play Between Worlds: Exploring Online Game Culture*. MIT Press, 2006.

[17] Turing, A.M. Computing Machinery and Intelligence. *Mind 59*, 236 (1950), 433-460.

[18] 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.

[19] 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.

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

[21] 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

[22] 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.

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