Medicine Meets Virtual Reality 17
J.D. Westwood et al. (Eds.)
IOS Press, 2009

doi:10.3233/978-1-58603-964-6-253

A Virtual Human Agent for Assessing Bias in Novice Therapists

Thomas D. PARSONS[a,1], Patrick KENNY[a], Louise COSAND[a], Arvind IYER[a], Chris COURTNEY[a], and Albert A. RIZZO[a]
[a] *University of Southern California*

Abstract. Monitoring the psychological and physiological activity of persons interacting with virtual humans poses exacting measurement challenges. Three experiments are reported in this paper. In these experiments we made use of Virtual Human Agent technology to assess persons' psychological and physiological responses to Virtual Standardized Patients. The first experiment provided support for the usability of the Virtual Standardized Patients through the use of a virtual character emulating an adolescent male with conduct disorder. In the second experiment we further developed the technology and aimed at assessing whether novice mental health clinicians could conduct an interview with a virtual character that emulates an adolescent female who has recently been physically traumatized. The third experiment looked at the usability of Virtual Standardized Patients for eliciting psychophysiological responses following exposure to virtual humans representing different ethnicities.

Keywords. psychophysiology, affective bias, virtual humans, virtual patients

Introduction

Research is needed that directly evaluates the contribution of particular forms of bias to disparities in the area of mental health care. The potentially biased views of clinicians in training can be held knowingly or unknowingly and can result in action or a failure to act. Bias occurs in the beliefs and actions of individual clinicians in training. Bias also occurs when the unfounded assumptions of training clinicians become normative beliefs shared by members of practitioner networks or treatment organizations.

1. Virtual Human Agents

The potential of using virtual humans as virtual standardized patients (VSP) for use in clinical assessments, interviews, and diagnostic training is becoming recognized as the technology advances. These VSPs are embodied interactive agents who are designed to simulate a particular clinical presentation of a patient with a high degree of consistency and realism [1]. VSPs have commonly been used to teach bedside competencies of bioethics, basic patient communication, interactive conversations, history taking, and

[1] Corresponding Author: Thomas D. Parsons, Ph.D., Research Scientist and Neuropsychologist, University of Southern California, Institute for Creative Technologies, 13274 Fiji Way, Marina del Rey, CA. 90292-4019, E-mail: tparsons@usc.edu.

clinical decision making. VSPs can provide valid, reliable, and applicable representations of live patients.

Research into the use of VSPs in psychotherapy training is in its nascent stages. Since virtual humans and virtual environments can allow for precise presentation and control of dynamic perceptual stimuli (visual, auditory, olfactory, gustatory, ambulatory, and haptic conditions), conversations and interactions, they can provide ecologically valid assessments that combine the control and rigor of laboratory measures with a verisimilitude that reflects real life situations. Although progress has been made toward establishing systems that are sensitive to component psychological processes, more studies are required to understand the effectiveness of these systems for training and education, to measure the believability of the characters with respect to their verbal and non-verbal behavior, and the ways in which differences in gender, ethnicity, and personality impact interactions with the VSPs.

2. Three Experiments using Virtual Standardized Patients

In this paper we describe a series of experiments with novice clinicians to evaluate the usefulness and effectiveness of VSPs as a medium to communicate with the students. The following description of the three experiments reveals the evolution of our VSP protocols. While initial assessment of the VSP was concerned with the system as a whole, later work emphasizes refined development and analysis of the human-computer dialectic (human questions and VSP responses). Additionally, as the VSP work has progressed, we are increasingly emphasizing investigation of the dialectic through the use of a number of psychological variables—openness to interaction with the VSP and willingness to be immersed in the virtual environment. Finally, we make use of psychophysiological metrics to assess affective bias.

2.1 Experiment 1: Virtual Standardized Patient "Justin": Adolescent Male with Conduct Disorder

The first project involved the construction of a natural language-capable VSP named "Justin." The clinical attributes of Justin were developed to emulate a conduct disorder profile as found in the Diagnostic and Statistical Manual of Mental Disorders (DSM IV-TR). Justin portrays a 16-year old male with conduct disorder who is being forced to participate in therapy by his family. Justin history is significant for a chronic pattern of antisocial behavior in which the basic rights of others and age-appropriate societal norms are violated. He has stolen, been truant, broken into someone's car, been cruel to animals, and initiated physical fights.

Assessment of the system was completed by 1) experimenter observation of the participants as they communicated with the VSP; and 2) questionnaires. To adequately evaluate the system, we determined a number of areas that needed to be addressed by the questionnaires: 1) The behavior of the VSP should match the behavior one would expect from a patient in such a condition (e.g. verbalization, gesture, posture, and appearance); 2) Adequacy of the communicative discourse between the VSP and the participants; 3) Proficiency (e.g. clarity, pace, utility) of VSPs discourse with the participant; and 4) Quality of the speech recognition of utterances spoken.

Although there were a number of positive results, the overall analysis revealed some shortcomings. First, positive results from the questionnaires were found related to

overall system performance. Further, participants reported that the system 1) simulated the real-life experience (i.e. ranked 5 or 6); and 2) the verbal and non-verbal behavior also ranked high (i.e. between 5 and 7). However, results also revealed that some participants found aspects of the experience "frustrating". For example, some participants complained that they were receiving anticipated responses and the system tended to repeat some responses too frequently. This was due to the speech recognition's inability to evaluate certain of the stimulus words. Further, there were too many "brush off" responses from the VSP when participant questions were outside the VSP's dialog set. There was also a concern that participants ascribe characteristics to the VSP which in fact are not present. For example, although the VSP responded "Yes" to a question about whether the VSP "hurt animals", in actuality the system did not recognize the input speech. This may lead to confusion if the VSP responds inconsistently. In fact one of the most substantial lessons learned was the amount of conversation tracking needed for the topical questions asked by the participant to allow the VSP's responses to be consistent throughout the session.

2.2 Experiment 2: Virtual Standardized Patient "Justina": Female Adolescent with Post-traumatic Stress Disorder

For Experiment 2, we built an adolescent female character called Justina that had been the victim of an assault and shows signs of PTSD. The technology used for the system is based on the virtual human technology developed at USC [2] and is the same as what was used with the previous VSP 'Justin'. To assess whether the responses of a VSP simulating PTSD in an adolescent female could elicit a number of DSM IV TR-specific questions (from novice clinicians) that are necessary for differential diagnosis, our data analysis was completed in two stages. In the first stage, the reference distribution is a correlation of each cluster of questions (from the novice clinicians) making up a particular DSM IV-TR PTSD Category with each (corresponding) cluster of responses from the VSP representing the same DSM IV-TR PTSD Category. In the second stage, variance from each individual's psychological distribution is controlled. Herein, the reference distribution reflects a semi-partial correlation controlling for the psychological factors that may be impacting the relation between each cluster of questions (from the novice clinicians) making up a particular DSM IV-TR PTSD Category with each (corresponding) cluster of responses from the VSP representing the same DSM IV-TR PTSD Category. We also assessed the impact of absorption and immersiveness upon the "believability" of the system.

Participants were asked to take part in a study of novice clinicians interacting with a VSP system. A total of 15 people (6 females, 9 males; mean age = 29.80, SD 3.67) took part in the study. The subject pool was made up of three groups: 1) Medical students (N=7); 2) Psychiatry Residents (N=4); 3) Psychiatry Fellows (N=4).

The primary goal in this study was evaluative: can a VSP generate responses that elicit user questions relevant for PTSD categorization? Findings suggest that the interactions between novice clinicians and the VSP resulted in a compatible dialectic in terms of rapport, discussion of the traumatic event, and the experience of intrusive recollections. Further, there appears to be a satisfactory amount of discussion related to the issue of avoidance. These results comport well with what one may expect from the VSP (Justina) system. A secondary goal was the investigation of the relationship between a number of psychological variables and the resulting VSP responses. A

summary of relations (measures as effect sizes "r") were generated between 1) each DSM IV-TR PTSD Category cluster of user questions; and 2) each (corresponding) cluster of responses from the VSP representing the same DSM PTSD Category. Please note that these are "clusters" of Question/Response pairs that reflect different diagnostic categories used for differential diagnosis. Moderate effects existed between User Questions and VSP Response pairs for "reporting of traumatic event" (r = 0.45), "re-experiencing" (r = 0.55), "avoidance" (r = 0.35), and "rapport" (r = 0.56), but only small effects were found for "hypervigilance" (r = 0.13) and "impairment of social" (r = 0.13). After controlling for the effects of the Tellegen Absorption Scale, increased effects were found for "reporting of traumatic event" (r = 0.48), "avoidance" (r = 0.37), "hypervigilance" (r = 0.15), and "impairment of social" (r = 0.24).

2.3 Experiment 3: Psychophysiological Assessment of Immersion using the Virtual Reality Cognitive Performance Assessment Test

In Experiment 3, we measured the activation and control of affective bias using 1) startle eye blink responses; and 2) self-reports as white participants interact with white and black virtual humans. By measuring eyeblink responses to startle probes occurring at short and long latencies following the onset of Black compared with White VSPs, we were able to examine affective processes associated with both the activation and potential control of bias.

Individual differences in levels of bias were predicted using E. A. Plant and P. G. Devine's (1998) Internal and External Motivation to Respond without Prejudice scales (IMS/EMS) [3]. Since participants with varying levels of IMS and EMS may differ in their responses to general affective stimuli unrelated to ethnicity, we presented general affective pictures from the International Affective Picture System (IAPS; Center for the Study of Emotion and Attention, 1995). Finally, we obtained participants' responses to the Attitude Toward Blacks scale to compare the predictive ability of both the virtual human interaction and the IMS–EMS with a traditional attitude measure of prejudice.

The initial participant pool included 14 adults at the University of Southern California. Eight were Caucasian, one was African-American, and five identified as other ethnicities. All participants read and signed an informed consent form. However, numerous subjects were excluded from the analysis due to the high threshold required to analyze blink amplitude. The resulting participant pool included five participants. Four of the five participants were Caucasian and one was African American. The African American participant responded similarly to Caucasian participants on physiological measures.

Eye blinks were collected and scored as electromygraphic (EMG) activity of the orbicularis oculi muscle of the left eye according to standard procedures (see Blumenthal et al., 2005; Biopac, Santa Barbara, CA). One small (4 mm) silver-silver chloride (Ag-AgCl) electrode was placed on the left eyelid directly below the pupil while a second 4 mm electrode was placed approximately 1 cm lateral to the first. The impedance between the two electrodes was measured and deemed acceptable if below 10 KΩ. A large (8 mm) Ag-AgCl electrode was placed behind the left ear to serve as a common ground. Startle blinks were identified in each portion of the recordings as follows: an interval (of 150 ms) adjacent to each startle probe was searched for spikes whose absolute value exceeds a high threshold (of 150 μV).

A difference score between the median blink amplitude to African American VSPs and Caucasian VSPs was determined. The difference score was then analyzed in terms of IMS and EMS scores. As all participants reported similar levels of internal motivation to respond without racism (Range = 6.2 – 7.4), they were separated into two groups of higher (4.0 and above, N = 2) and lower (below 4.0, N = 3) external motivation scores. A one-way ANOVA was performed to determine if high vs. low external motivation related to physiological responses to different racial groups of virtual patients. The difference score tended to be lower (M = 0.007; SD = 0.013) in those with high external motivation to behave in a non-racist manner. Those who were lower in external motivation had a larger difference score (M = 0.020; SD = 0.007) between startle amplitudes while looking at African American vs. Caucasian VSPs. The larger difference score reflects larger startle amplitudes to African American VSPs, suggesting an implicit negative bias towards that group. The difference between the low EMS and high EMS groups was not significant ($F= 2.30$; $p = 0.23$) but suggests need for future analyses into the ways in which motivations can influence behavior at even automatic levels.

3. Discussion

In this paper we described a series of experiments with novice clinicians to evaluate the usefulness and effectiveness of VSPs as a medium to communicate with the students. The evaluation consisted of an assessment of the system as a whole through questionnaires and data collection of the questions and responses in the interview. Additionally we investigated the relationship of the questions with a number of psychological variables such as openness to interaction with the VSP and willingness to be immersed in the virtual environment. Finally, we make use of psychophysiological metrics to assess affective bias.

The use of VSPs provides a window into clinicians' implicit biases, permitting a more thorough understanding of how clinicians evaluate and interact with patients. The patients permit a training opportunity in which ethnicity and gender sensitivity can be measured and modified. Clinicians are unlikely to express overt racist tendencies, but they may harbor tendencies of which they are embarrassed or even unaware. Self-reported racism has declined over the last several decades [3], yet a significant factor in the reported decrease may be due to changing social standards' effects on self-report data. Self-report data are susceptible to modification by a participant's awareness of the social desirability of particular responses [4], reducing the sensitivity of the measures. Implicit behavioral and psychophysiological responses, however, are automatic and thus considered less susceptible to self-conscious influences [5].

While the first two experiments reflect the development and assessment of VSPs for communication competencies, the third experiment used electromyographic responses to acoustic startles and self-report data to better understand how participants relate to VSPs of different ethnicities. The results suggested a minor trend towards an additive effect of sources of motivation on participants' physiological responses to VSPs of different ethnicities. All participants surveyed reported high internal motivation to behave in a non-racist manner. Participants differed on their levels of external motivation to suppress racist tendencies, and those who were high on external motivation showed more similar blink amplitudes to black and white VSPs than did those participants reporting low external motivation. This effect on physiological

responses may seem counter-intuitive, as external motivations are consciously processed and generally not thought to influence automatic processes [5]. These data suggest that automatic processes are not completely impervious to modulation by the conscious self and the additive nature of internal and external motivation may result in a stronger suppression of racist reactions. Humans are social animals and logic suggests that it would be adaptive to allow social norms to shape even automatic processes.

In sum, these three experiments represent early attempts to enlist VSPs in aiding our understanding of communication skills and potential biases in novice clinicians. We believe that it is useful to consider the regulation of bias along a continuum of responses, varying in controllability from the most explicit to the most implicit. Our hope is that, by identifying individual differences in training clinician's regulatory effectiveness for both explicit and implicit expressions of bias, we can enhance our understanding of the steps that must be taken in the bias reduction process. The present conceptualization would allow us to identify training clinicians at different stages in the bias reduction process, marked by the sources of motivation impelling them to regulate expressions of bias.

References

1. Kenny, P., Parsons, T.D., Gratch, J., Leuski, A., & Rizzo, A.A. (2007). Virtual Patients for Clinical Therapist Skills Training. *Lecture Notes in Artificial Intelligence* **4722** (2007).
2. Swartout, W., Gratch, J., Hill, R., Hovy, E., Marsella, S., Rickel J., Traum, D.: Toward Virtual Humans, AI Magazine, **27**(2006).
3. Plant, A.E., & Devine, P.G. Internal and external motivation to respond without prejudice. *Journal of Personality and Social* Psychology **75** (1998), 811-832.
4. Schwarz, N. Self-reports: How the questions shape the answers. *American Psychologist* **54** (1999), 93-105.
5. Phelps, E.A., O'Connor, K.J., Cunningham, W.A., Funayama, E.S., Gatenby, J.C., Gore, J.C., & Banaji, M.R. Performance on indirect measures of race evaluation predicts amygdala activation. *Journal of Cognitive Neuroscience* **12** (2000), 729-738.