Rethinking some Virtual Human Applications

Robert HUBAL[a,1]

[a]*Renaissance Computing Institute, Chapel Hill, NC*
ORCiD ID: Robert Hubal https://orcid.org/0000-0003-2637-0918

Abstract. Increasingly realistic virtual environments incorporating virtual characters have been used to train or assess actual behavior, such as of people at risk, and identify reasons to remediate or intervene. Technology has improved so rapidly that today's capabilities to create situations to focus training and intervention outshine past efforts. To name just a few current examples, tools like Unreal's MetaHuman Creator for creating characters, Midjourney for creating environments, OpenAI's ChatGPT for scripting, and GIFT for tutoring have enormous potential, as these tools promise to reduce simulation costs and increase realism. This paper, in contrast, discusses some movement in the other direction: Recent efforts suggest that increased realism may not always have resulting cost- benefit for training and assessment. Lessons learned and recommendations are presented to guide future developers.

Keywords. Consequential conversations, responsive virtual humans, character and application fidelity.

1. Introduction

Social skills deficits show up as poor behaviors in demanding social situations. For example, in a confrontational situation, some individuals misperceive linguistic, gestural, and expressive cues of another person as indicating hostile intent. Misinterpretation of cues could lead to an escalation of the situation. 'Consequential' conversations, as used here, are those that involve challenging content, may lead to adverse outcomes, and require deft social interaction skills to navigate. The partner(s)in the conversation may be difficult to deal with, emotional or confused, or focused on an agenda. The topic of conversation may be sensitive, charged, controversial, or zero- sum.

These situations are uncomfortably common. Military personnel returning from stressful combat demonstrate social skills deficits in numerous ways including domestic discord, substance use, and generalized aggression and violence [1,2]. Law enforcement personnel who are not trained in crowd control can and do misunderstanding actions taken by protesters and, rather than use de-escalation skills, apply unnecessary force [3,4]. Conversely, minority citizens may perceive bias from authority figures [5] even when unintended and speak angrily and act out unsafely. More mundanely, therapy targeted to individuals who demonstrate social skills deficits (e.g., those with some forms of autism, or depression or other mental health concerns) is often general, given measurement techniques [6], whereas their deficits are specific. In all of these scenarios, training or education or treatment should instead be targeted to those behaviors that pose greatest risk to the individual (or those around the individual) and thus greatest benefit when addressed. Typical assessments of social skills ask individuals what they would do in hypothetical situations, or they present non- interactive situations to gauge individuals'

[1] Corresponding author: Robert Hubal, hubal@renci.org

reactions, or they present questionnaires. These typical assessments are thus descriptive—not demonstrative—in nature.

1.1. Virtual Human Applications

A different approach is to use a gaming engine and established artificial intelligence (AI) models to present situations to gauge behavior. A key affordance of game-based technology is the capacity to provide users with an opportunity to role-play identities inthe context of a realistic scenario [7]. The virtual environment provides opportunities for users to participate as well as vicariously observe virtual characters in interacting with the situation. Here, the idea is to elicit decision making and social skills, including underlying abilities such as emotional control, information seeking, expressing preferences, negotiation and willingness to compromise, and using non-provocative language. The skills are assessed by simulating social encounters that may lead to adverse consequences. In these applications, individuals' behavior is assessed by observing them while they interact within the simulated environment (e.g., looking at body language, tone of voice, emotional control, and reaction time) and by looking at what type of outcome occurs—usually a good outcome demonstrating avoidance of risky behavior or a poorer one demonstrating risky behavior.

Software applications designed to train or assess dialog within consequential conversations have, for many years, employed virtual characters. Virtual characters are, paradigmatically, multimodal embodied conversational agents—responsive partners with which a user communicates to navigate a given situation or achieve a goal. Virtual character applications have various advantages, including their reproducibility, safety and controllability, ease of distribution, and objectivity, and the ability to introduce intelligent tutoring. Typically, such applications have engaged a user using realistic virtual participants in a realistic setting. Multiple situations address a range of social skills competencies, including poor emotion expression recognition, impulsivity, insensitivity to penalties, hostility bias, gender stereotyping, acceptance of dating or partner violence, and risky decision making. The virtual environments are designed to be flexible and reusable, so that, by having runtime parameters define any givensituation, minor visual (e.g., the characters' appearance) or behavioral (e.g., a character acts angry vs. confused) variation is possible, or major changes, and situations can be repurposed.

2. Rethinking Realism

Realism has increased dramatically as technology and capability have improved, so that today's characters can be made to be lifelike in appearance, allow for natural language interaction, and use advanced behavior models to react or respond to user actions appropriately to the context. Over the years this author [8-12] and many others have used evolving tools to portray simulated situations. This author started out in the 1990s with a homegrown renderer built atop DirectX; nowadays established game engines such as Unreal and Unity provide more than the functionality needed. IBM's ViaVoice was once used to capture language input augmented by custom grammar and contexts to capture state of the dialog. Now, dialog systems for consequential applications can take advantage of large language models. For this author, the culmination of work was a virtual patient represented in **Figure 1** developed for pharmacy students to learn skills associated with medication therapy management [13].

Figure 1. Virtual patient.

2.1. Some Lower-Fidelity Examples

Unexpectedly, perhaps, then, this researcher has moved toward lower fidelity in some recent applications. The shift has been neither total nor abrupt, but, in retrospect, unforeseen.

One example of a less realistic consequential virtual human application is represented by a primarily text-based interactive simulation to present pharmacy preceptors with challenging non-academic situations [14]. There were several reasons for using a web-based, low-interactive approach using Ink as an underlying dialog control mechanism. First, the study was conducted during the Covid pandemic, requiring virtual participation. Distribution of a game-based application might have been difficult. Further, because of the means of engagement (virtual via Zoom), a true natural language interaction, though possible, was infeasible, suggesting that a high- fidelity virtual character also was not needed. Instead, as evidenced by their think-aloud,the situations themselves engaged participants sufficiently. While specifics of the study would be different for any other, these types of determinants may apply elsewhere.

Another example is a branching video application depicted in **Figure 2** for use in helping military personnel understand good techniques engaging foreign civilians [10]. Again, there were a couple of reasons for this choice of technology. First is the availability of resources; the project was better able to support a less immersive than a more immersive approach. Second was the capability of the technology. Military personnel were asked not only to engage with presented characters but also to mark up important or concerning elements of the presented scene, and the technology supported this requirement. Note that the conversations themselves were no less consequential than others where responsive virtual human technology was used. Instead, it was the engagement with presented situations that supported the learning desired.

Another example (in which the author was only indirectly involved) is a graphic novel shown in **Figure 3** developed to illustrate, also for military personnel, life in a combat zone [15]. The approach was chosen to target a younger audience and provide realistic enough situations, "embedding the principles of combat and operational stress control…into realistic and relatable characters, stories, and images" [16], that would prepare personnel for deployment. The author and colleagues have recently worked on a similar design for a parenting application directed toward, and to appeal to, those recovering from substance abuse.

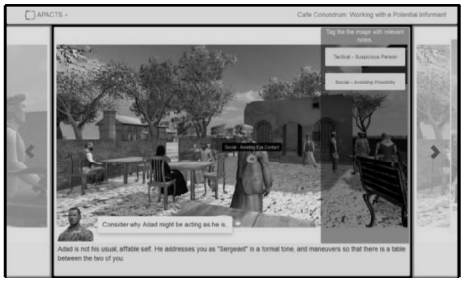

Figure 2. Engaging foreign civilians.

Figure 3. Combat.

A final example is a virtual reality intervention that allows hospitalized adolescents to learn and practice therapeutic skills in stressful situations [17]. At first glance an immersive application does not appear to be low fidelity, but in fact the visuals are necessarily limited due to resource constraints and the dialog interaction is controlled. However, the situations themselves have been heavily tested and shown to be reflective of real events, and adolescents demonstrate engagement within the situations despite noting a perceived lack of realism.

2.2. Lessons Learned and Recommendations

The intentions of this paper are to challenge a presumption that high fidelity virtual characters and environments are always the best solution for training or assessment in consequential conversations, and to encourage further research. It is not to discourage greater fidelity when demanded.

When is greater fidelity demanded? One example is when expressions, and in particular changes in expression, need to be noticed, as when there are pressing issues such as interaction or cultural nuances that need to be made explicit. Branching video would be a possibility here [18], though video is not usually as flexible as virtual settings that can easily support a swap of characters of variable ages, ethnicities, and other characteristics, so that the user experiences a range of characters. Relatedly, usage of more realistic virtual characters makes sense when the user needs to engage meaningfully with the character as in talk therapy. Also, higher fidelity is called for when assessing fine-grained procedures that would otherwise be logistically complex, unsafe, unethical, unreliable, and/or costly. For instance, programs where medical personnel need to demonstrate precisely where to apply specific instruments benefit from realistic characters [19].

When is lesser fidelity acceptable? Virtual character realism makes less sense when the application does not involve interactivity, such as when characters are in the background or at least not central to the task at hand. When an application centers on physical manipulation, including haptic feedback, then other methods such as manikins or part-task trainers, or a hybrid approach [20], are typically more cost-effective.Further, when users engage just as fully with less realistic characters as more realistic characters, then the former represent a more cost-effective training or assessment.

The author's change in approach for some—not all—applications derives from several additional conditions: Difficulty in developing suitable models to meet user expectations, resource and usability constraints, learner preferences, reappraisal of the purpose of training or assessment, and affordances of underlying technology. For instance, behavior models to drive character actions may need to be highly refined in applications that show nuance (e.g., subtle medical symptoms) but coarser in others where a general appearance of motion (e.g., with graphic novels) is sufficient. Similarly, users can be pretty much assured to be engaged in applications portraying well-designed consequential conversations, particularly when they relate to their lives, and barring an egregious error.[2] When that engagement demands higher fidelity depends on factors including learning objectives, costs, distribution, and additional technological needs. Also, use of current technologies such as those mentioned above that facilitate

the development and portrayal of realistic characters leads users to expect realistic behaviors, akin to the uncanny valley effect [21].

This author addresses the following questions to guide development: What is the responsibility of users, and what are their actions? Are any skills to be gained meant to be familiarized, practiced, or mastered (the former may need lesser engagement than the latter)? Are the skills procedural, interactive, or strategic (the latter may need greater engagement than the former)? Is the task dangerous, complex, time-consuming, resource-consuming, expensive, or unavailable, so that a realistic simulation promises benefits in relation to costs? How easy or hard is it to model behaviors of entities and objects (the harder to model, the less the fidelity is probably called for)? How visually demanding is the task (the more demanding, the greater fidelity likely required)?

These criteria apply differently for any given application but represent principles toconsider as the application is designed. Rapidly advancing technology is exciting and generally worth taking advantage of. But getting users to engage is the lesser challenge; the greater challenge is to neither under- or over-promise what they experience.

References

[1] Kline AC, Panza KE, Nichter B, Tsai J, Harpaz-Rotem I, Norman SB, Pietrzak RH. Mental health care use among U.S. military veterans: Results from the 2019–2020 National Health and Resilience in Veterans Study. Psychiatr Serv. 2022;73(6):628-35. doi: https://doi.org/10.1176/appi.ps.202100112

[2] MacManus D, Rona R, Dickson H, Somaini G, Fear N, Wessely S. Aggressive and violent behavior among military personnel deployed to Iraq and Afghanistan: Prevalence and link with deployment and combat exposure. Epidemiol Rev. 2015:37(1):196-212. doi: https://doi.org/10.1093/epirev/mxu006

[3] Engel RS, McManus HD, Herold TD. Does de-escalation training work? A systematic review and call for evidence in police use-of-force reform. Criminol Public Pol. 2020;19(3):721-59. doi: https://doi.org/10.1111/1745- 9133.12467

[4] Trinkner R, Kerrison EM, Goff PA. The force of fear: Police stereotype threat, self-legitimacy, and support for excessive force. Law Hum Behav. 2019;43(5):421-35. doi: https://doi.org/10.1037/lhb0000339

[5] Swencionis JK, Goff PA. The psychological science of racial bias and policing. Psychol Public Policy Law. 2017;23(4):398-409. doi: https://doi.org/10.1037/law0000130

[6] Cordier R, Speyer R, Chen YW, Wilkes-Gillan S, Brown T, Bourke-Taylor H, Doma K, Leicht A. Evaluating the psychometric quality of social skills measures: A systematic review. PLoS One. 2015;10(7):e0132299. doi: https://doi.org/10.1371/journal.pone.0132299

[7] Bachen CM, Hernández-Ramos P, Raphael C, Waldron A. How do presence, flow, and character identification affect players' empathy and interest in learning from a serious computer game? Comput Hum Behav. 2016;64:77-87. doi: https://doi.org/10.1016/j.chb.2016.06.043

[8] Hourani LL, Kizakevich PN, Hubal R, Spira J, Strange LB, Holiday DB. Predeployment stress inoculation training for primary prevention of combat-related stress disorders. J Cyberther Rehabil. 2011;4(1):101-16.

[9] Hubal RC, Evens NR, FitzGerald DP, Hardy KK, Willard VW, Bonner MJ. Implementation of FACS for synthetic characters for use in studying facial expression recognition by survivors of childhood cancer. Annu Rev Cyberther Telemed. 2008;6:35-40.

[10] Hubal RC, Kizakevich PN, Guinn CI, Merino KD, West SL. The virtual standardized patient – Simulated patient-practitioner dialogue for patient interview training. Stud Health Technol Inform. 2000;70:133-8. doi: https://doi.org/10.3233/978-1-60750-914-1-133

[11] Hubal R, van Lent M, Wender J, Lande B, Flanagan S, Quinn S. What does it take to train a good stranger. Procedia Manuf. 2015;3:3955-62. doi: https://doi.org/10.1016/j.promfg.2015.07.930

[12] Paschall MJ, Fishbein DH, Hubal RC, Eldreth D. Psychometric properties of virtual reality vignette performance measures: A novel approach for assessing adolescents' social competency skills. Health Educ Res. 2005;20(1):61-70. doi: https://doi.org/10.1093/her/cyg103

[13] Hubal R, Heneghan J. Carolina virtual patient initiative [abstract]. Pharm Educ. 2017;17(1):292.

[14] Williams CR, Hubal R, Wolcott MD, Kruse A. Interactive narrative simulation as a method for preceptor development. Pharm. 2021;10(1):5. doi: https://doi.org/10.3390/pharmacy10010005

[15] Kraft H, Peeler R, Laron J, Borst CA, Hourani LL, Lambert S, Wiggins D, Nguyen D, Philyaw D, Jessup E. The docs: A graphic novel. San Diego: Naval Health Research Center; 2010.

[16] Hourani L, Lambert S, Peeler R, Lane B, Borst C. Graphic novels: A new stress mitigation tool for military training: Developing content for hard-to-reach audiences. Health Commun. 2017;32(5):541-9. doi: https://doi.org/10.1080/10410236.2016.1140265

[17] Marraccini M, Hubal R, Middleton T, Delgaty L. Development and testing of a virtual-reality intervention with youth hospitalized for suicide-related crises. Annu Rev Cyberther Telemed. in press.

[18] Smith MJ, Bornheimer LA, Li J, Blajeski S, Hiltz B, Fischer DJ, Check K, Ruffolo M. Computerized clinical training simulations with virtual clients abusing alcohol: Initial feasibility, acceptability, and effectiveness. Clin Soc Work J. 2021;49(2):184-196. doi: https://doi.org/10.1007/s10615-020-00779-4

[19] Kizakevich PN, McCartney ML, Nissman DB, Starko K, Smith NT. Virtual medical trainer - Patient assessment and trauma care simulator. Stud Health Technol Inform. 1998;50:309-15. doi: https://doi.org/10.3233/978- 1-60750-894-6-309

[20] Brown WJ, Tortorella RAW. Hybrid medical simulation – A systematic literature review. Smart Learn Environ. 2020;7:16. doi: https://doi.org/10.1186/s40561-020-00127-6
[21] Wang S, Lilienfeld SO, Rochat P. The uncanny valley: Existence and explanations. Rev Gen Psychol. 2015;19(4):393-407. doi: https://doi.org/10.1037/gpr0000056